Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

The Generalization of Attitude Change within a Serial Structure: H. PEAK, H. W. MORRISON, AND R. P. QUINN 281

Transfer of a Discrimination from Objects to Patterns: B. J. HOUSE AND D. ZEAMAN 298

Eyelid Conditioning as a Function of Intensity of Conditioned and Unconditioned Stimuli: E. G. WALKER 303

Tracking Performance on a Sequence of Step Functions which Approaches a Continuous Function as a Limit: D. MCCONNELL AND M. W. SHELLY 312

Multidimensional Mapping of Munsell Colors Varying in Hue and Chroma: T. INDOW AND T. UCHIZONO 321

Multidimensional Mapping of Munsell Colors Varying in Hue, Chroma, and Value: T. INDOW AND K. KANAZAWA 330

Resistance to Extinction as a Function of Degree of Reproduction of Training Conditions: M. H. MARX 337

Replication Report: An Attempt to Obtain Inhibition with Reinforcement: S. S. FUCHS 343

American Psychological Association

Vol. 59, No. 5 May 1960

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.*** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1960 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
BEATRICE B. HAMLIN, Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1960 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 59, NO. 5 MAY 1960

THE GENERALIZATION OF ATTITUDE CHANGE WITHIN A SERIAL STRUCTURE

HELEN PEAK, H. WILLIAM MORRISON, AND R. P. QUINN[1]

University of Michigan

This study is one of a series devoted to exploring the generalization of change in attitude or affective responses within different psychological structures. It has often been assumed that antecedent–consequent patterns, such as are characteristic of a serial structure, are involved in so-called secondary or mediated generalization, and one of the unsettled issues about mediated generalization is the question of the relative importance of factors operating at the time of response training and at the time of testing for generalization (Razran, 1949).

If every act depends on the combined effects of a specific input pattern (from external and internal sources) activating an existing structure, then the response made to any stimulus being tested for generalization will be influenced by these factors. Consequently, anything which at the time of response training alters the psychological structure of which the stimulus-to-be-tested is a part will influence the appearance of generalization. Likewise any alteration in the total input to that structure at the time of testing may alter the probability of the occurrence of a particular generalized response.

Thus if a stimulus which is part of a previously learned serial structure is presented and paired with a response, it is likely to elicit with some frequency each other "stimulus" which has been learned as part of the same series. As a consequence, the response following the stimulus-in-training will become associated with the other stimuli as well. An alteration in associative structure will have occurred as a result of response training and generalization will be affected in the manner to be explained. Moreover, when a test stimulus which is part of such a serial structure is presented for rating, the event which is most likely to be activated is the forward associate of the test stimulus.

[1] This investigation was carried out under a Contract between the Office of Naval Research and the University of Michigan, Project Nonr 1224(10)NR171-039. The senior author was in residence at the Center for Advanced Study in the Behavioral Sciences and the second author was a Public Health Service Research Fellow of the National Institute of Mental Health when the article was written. Margaret Clay and Vera Schlesinger helped in the collection of part of the data.

It will be suggested that this and other events interact to determine the result at the time of testing for generalization.

In the experiment reported here a serial structure was established by rote learning of 12 nonsense syllables during relations training. After completion of this phase of training the seventh stimulus in the series (S_7) was associated with shock during response training. At the beginning and end of the experiment *Ss* rated syllables S_4 through S_{10} on scales so that change in the judged affect of the trained stimulus (S_7) and six test stimuli could be measured. In a Paired-Associates Control (PA) condition, instead of learning the serial list, *Ss* learned the same syllables as responses to numbers; other procedures were the same as in the serial condition.

Theory and Hypotheses

The serial structure.—In the present experiment generalization of response based on processes occurring either during response training or during tests for generalization depends on the nature of the serial structure. It is well known that this structure is not a simple sequence of related units but rather a complex of backward and forward associations of different degrees of remoteness

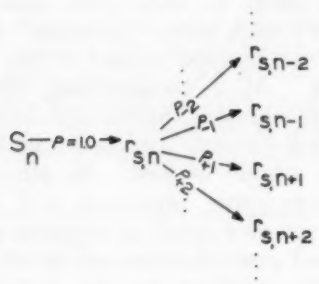

Fig. 1. Activation of units in a serial structure when one stimulus in the *n*th position is presented. Subscripted *p* indicates the probability of activation of units in each list position at a specified number of steps from *n*.

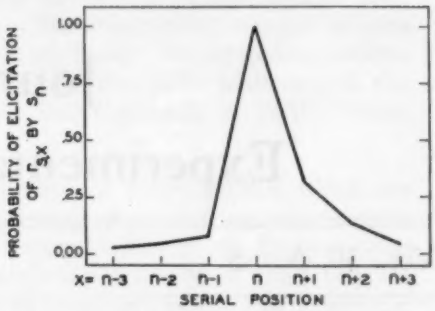

Fig. 2. Probability of activation of units in a serial structure when the *n*th stimulus is presented (estimated from Raskin and Cook, 1937).

Whenever some stimulus (S_n) within the series is presented, it is assumed that an implicit event ($r_{s,n}$) is always elicited. It is further assumed that following the elicitation of $r_{s,n}$ there is some probability of the elicitation of each other unit in the series ($r_{s,1}$ to $r_{s,12}$). These events are represented in Fig. 1, which omits activation of r_s's from sources other than $r_{s,n}$. The probabilities of the activation of these events forward and backward from some one point in a series may be approximated from the Raskin and Cook study (1937), which reports the frequencies with which serially associated nonsense syllables in various list positions were given as associates to syllables in other positions. Figure 2 represents the relative frequency of response observed at various list positions relative to the position of the stimulus syllable.[2] In the PA Control

[2] This study has been criticized because of methodological imperfections (Bugelski, 1950). It should be noted, however, that for our purposes the fact that the same nonsense syllable was always used in the same serial position does not raise the problems to which Bugelski refers. Every stimulus but the last was the point of origin for a one-step forward associate and every stimulus but the first was the origin for a one-step backward associate, so the frequencies at a given position relative to the stimulus syllable were not dependent on responses to one stimulus. Moreover, their technique of asking for free associations to each syllable as a means of determining its probability of occurrence was a more appropriate method for ascertaining the relative frequencies of response than the Bugelski method of counting intrusions in the course of learning by the anticipation method.

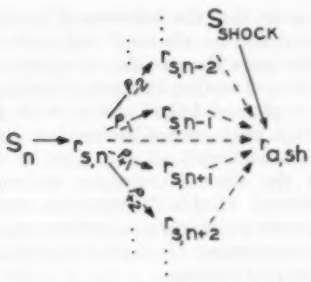

FIG. 3. The association of serial units with the affective response to shock ($r_{a,sh}$) when S_7 (S_n) is presented. A solid line indicates a previously learned association. A broken line indicates the formation of a new association at the time of response training. The contiguous occurrence of any r_s and $r_{a,sh}$ results in learning an affective response to that r_s.

condition, on the other hand, it is assumed that such relations between nonsense syllables are weak and equal in strength.

Effects of response training.—From the results described above, it may be inferred that when S_7 is presented and followed by shock, each other unit of the serial structure has some probability of elicitation and of being associated with shock. This probability depends on the list position of a given unit relative to S_7. Although never presented during response training, these implicit "stimulus" events as well as the trained stimulus will be contiguous with the negative affective reaction ($r_{a,sh}$) elicited by the shock. The frequency of association of each unit with $r_{a,sh}$ is one determinant of the affective judgment of that stimulus when later tested for generalization. Figure 3 represents the association of serial units with shock during training and the strength of negative affective response to each unit at the end of response training.

We have made a straightforward use of the Raskin and Cook data in estimating the relative strength of affective responses to each serial unit, assuming that relative response strength is a bilinear function of (*a*) relative frequency of elicitation of a given unit by S_7 and its consequent association with shock, and (*b*) the amount of change in the affective response to the trained stimulus as the result of pairing with shock. The results of such estimates are presented in Fig. 4. Three levels of trained change are illustrated. In a maximum change group, change in S_7 is given the arbitrary value of −1.00; also illustrated is a moderate change group with S_7 change represented as 50% of S_7 maximum (i.e., −0.5) and a minimum change group with S_7 equal to 25% of S_7 maximum (i.e., −0.25). Obviously, for different *Ss* change in S_7 may be any proportion of the maximum change value for S_7. In order to estimate the relative affective reaction to stimuli in each serial position in groups with S_7 change 100, 50, and 25% of S_7 maximum change, we have multiplied such values of S_7 by the probability of elicitation of any $r_{a,s}$ by S_7 taken from Fig. 2. Because of the shape of the curve in Fig. 2, the affective reaction to S_8, the forward associate of the trained stimulus, is more negative than that to any other untrained stimulus and there is a gradient falling off in both directions from S_7.[3]

On the basis of this analysis Hypotheses 1 and 2 were proposed:

1. The amount of negative change in the rating of S_8 decreases with decrease in the amount of negative change in S_7.

2. When the rating of S_7 changes most negatively, the rating of S_8 changes (*a*) more negatively than ratings of untrained stimuli in the PA Control condition, and (*b*) more negatively than ratings of syllables at all other serial positions except S_7.

Consideration of processes occurring at the time of testing do not affect the above predictions about generalization of response to S_8 (see below).

Effects at the time of test for generalization.—After one-third of the cases had been run it became apparent that our predictions were being fulfilled in that change in the rating of S_8 was more negative than that of other syllables in the list and more negative than PA Controls. However, it was also clear that the change in rating of S_8 did not conform to the prediction made on the basis of events during response training; namely that change in the rating of S_8 would be negative though not as negative as S_7 (see Fig. 4). Nor was the result the one which should have been predicted by Razran (1949) or by other proponents of a mediation theory (Osgood, 1953); presumably they would predict that the stimulus antecedent to a trained stimulus in a serial list would produce the most negative generalized response. Compared with appropriate controls the rating of S_6 changed positively in the group which showed maximum negative change in S_7. There was a less striking tendency for

[3] There is a question about the probabilities of arousal of syllables at the ends of the serial list but these are not involved here.

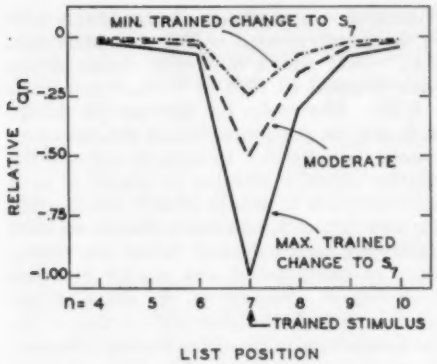

FIG. 4. Relative strength of implicit affective response to stimuli in different serial positions at three levels of change in response to the trained stimulus.

the rating of S_8 to change negatively with moderate negative change in the rating of S_7.

About the time that these results were obtained, Mink (1957) reported an experiment and a theory suggesting that the appearance of a generalized response is determined in part by the interaction between events occurring at the time of testing. Using paired word stimuli with much stronger associative strength from the first to a second word than in the reverse direction, he observed generalization of a voluntary finger response from the first to the second word, and compared this with generalization when the response was associated with the second word. The former arrangement yielded significantly more generalization. In the latter arrangement Mink attributed the absence of generalization in the direction of the trained response to *S*'s opportunity to compare the (first) test stimulus with the implicitly elicited trained "stimulus." Mink suggested that when *S* compared the tested and trained words he would realize that they were different and therefore would not respond to the test stimulus even though there was some tendency to do so.

One of us had previously elaborated a theory of activation within various kinds of structures which developed the notion that the occurrence of assimilation and contrast is a function of the psychological distance between two events,[4] and it became apparent at this point that the behavior of S_8 might be accounted for by contrast and assimilation effects between the complex of events elicited at the time of testing for generalization. The theory explained below shares with Mink's the notion that generalization is a function of events occurring at the time of testing and at the time of response training. It goes beyond Mink's formulation chiefly in stating more explicitly some of the conditions and consequences of such interactions between elicited events.

It is assumed that in our experiment the evaluative rating given to any syllable will be the result of the interaction of the various tendencies to make affective responses of some sign and intensity to the test syllable and to the other events activated on such an occasion. Again the properties of a serial structure are critical, for when a stimulus is presented for rating, the probability of elicitation of other serial units (and hence of their affective response tendencies) is determined as in Fig. 2. It should be noted that when S_n is presented, $r_{s,n+1}$, its forward associate, has a much higher probability of elicitation than other units in the list. In deriving predictions from processes at time of testing for generalization we do not consider elicitation of list units other than $r_{s,n+1}$ and $r_{s,n}$ by S_n. Figure 5 illustrates this simplified schema of the situation at the time of testing.

The test syllable, S_n, elicits $r_{s,n}$ and its associated effective response ($r_{a,n}$). At the same time the next serial list unit ($r_{s,n+1}$) has a high probability of being elicited, and will itself produce an affective response ($r_{a,n+1}$). The affective responses associated with each serial unit are those resulting from response training (see Fig. 4). The two levels of affective response ($r_{a,n}$) and ($r_{a,n+1}$) together with the input from the instructions to judge S_n, will interact to determine the overt affective judgment.

As already suggested, this judgment is assumed to be a function of the difference or distance between $r_{s,n}$ and $r_{s,n+1}$. Where there is a small absolute difference between

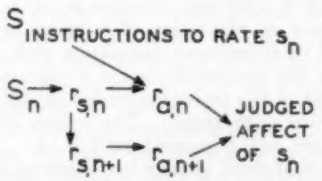

FIG. 5. Events at Time 2 rating of stimulus (S_n).

[4] This theory was discussed at the ONR Symposium at Harvard in March 1957. A revision of the paper presented at that time appeared in Tagiuri and Petrullo (Peak, 1958a). The theory is developed further in Peak (1958b).

these values, assimilation will occur and the judged affect of S_n is closer to that value corresponding to $r_{a,n+1}$ than if $r_{a,n+1}$ had not been elicited. At greater distances, contrast will result. In this case, the judged affect of S_n will be reduced when $r_{a,n} < r_{a,n+1}$ and increased if $r_{a,n} > r_{a,n+1}$, compared with the situation where $r_{a,n+1}$ is not elicited. Since the derivation of difference in affective responses for different list positions is relative, and since the absolute distances between events which result in assimilation or contrast are not known, predictions about these effects must be made in relative terms.

Following this paradigm, S_6 as a test stimulus will elicit not only its own affective response ($r_{a,6}$) but also its forward associate ($r_{a,7}$) which in turn produces $r_{a,7}$. Since the difference between $r_{a,6}$ and $r_{a,7}$ is greater than that for any other adjacent list positions (see Fig. 4), contrast is most likely to occur when S_6 is the test stimulus being rated. Furthermore, the more negative is $r_{a,7}$, the more likely it is that this difference will become great enough to produce contrast, since the difference in level of $r_{a,7}$'s between trained change groups is greater than the corresponding difference in $r_{a,6}$'s (see Fig. 4). The effect of such contrast between $r_{a,6}$ and $r_{a,7}$ will be to make the judgment of S_6 more positive.

Hypothesis 3 follows from these considerations: When negative change in affective rating of S_7 is greatest, the affective rating of S_6 will change (*a*) more positively than untrained syllables in the PA Control condition, (*b*) more positively than other syllables in the serial list, and (*c*) at the most negative level of trained change the rating of S_6 will change more positively than the rating of S_6 when there is less change in S_7.

Further consideration of the relations between $r_{a,6}$ and $r_{a,7}$ as $r_{a,7}$ becomes less negative suggests that when this difference decreases below the contrast level, negative change in the rating of S_6 will result, producing a more negative change in S_6 than in any other syllable except S_7 and S_8. This follows from the fact that S_4, S_5 and any other syllables antecedent to S_6 will activate r_a's of smaller magnitude than does S_6 (see Fig. 4), so that even if one or the other of these r_a's is assimilated to its forward associate (viz., $r_{a,5}$ or $r_{a,6}$) the judgment of S_4 or of S_5 would not be as negative as would the judgment of S_6 which would assimilate to the more negative $r_{a,7}$. Likewise, syllables following S_8 will produce smaller r_a's than does S_6, for if assimilation occurs to the forward associate of either S_9 or of S_{10} this will make the judgment of each still less negative. This leads to

Hypothesis 4: As change in negative affective rating of the trained stimulus decreases, change in the rating of S_6 will be (*a*) more negative than change in ratings of untrained syllables in the PA Control condition, and (*b*) more negative than change in any other stimuli in the serial list except S_7 and S_8.

Considering the effects of events at the time of testing on Hypotheses 1 and 2, it is clear that as $r_{a,7}$ becomes more negative, the distance between $r_{a,8}$ and $r_{a,9}$, its forward associate, increases. This would lead to increased contrast or decreased assimilation depending on the distance. Since this would have the effect of making the judgment of S_8 still more negative, Hypotheses 1 and 2, which are concerned with the greater negative change in S_8 when S_7 change is maximum, do not need modification.

Finally, the possible influence of these interaction processes on the judgment of S_7 should be mentioned. Figure 4 indicates that as $r_{a,7}$ becomes more negative the difference between $r_{a,7}$ and $r_{i,8}$ will be greater. This will increase the contrast between them, or decrease the assimilation which would make judgment of S_7 even more negative. It is clear, therefore, that judgments of S_7 would continue to reflect the greater negativity of $r_{a,7}$ as compared with other r_a's in spite of interaction of events at the time of rating.

Method

Subjects

Random assignment of 175 undergraduate women was made to the following conditions: Serial (72 *Ss*), Paired-Associates Control (73), and Test–Retest Control (30).

Design and Procedure

Table 1 presents the major features of the design and procedure. The *Ss* in each of the three different experimental conditions performed the tasks represented in the table and described below.

Affective ratings: Time 1.—All *Ss* rated 11 stimuli (2 meaningful words and 9 nonsense syllables chosen to minimize the number of common letters) on 12 semantic differential scales (Osgood, Suci, & Tannenbaum, 1957). The stimuli rated were FOG, STEEL, FAM, VED, CEF (always rated in Positions 1, 2, 3, 5, and 11, respectively, and ZUT, WIX, ROJ, YOF, BIP, and KAQ, whose rating positions varied within each group of six *Ss* so that each appeared equally often in each rating position. In the Test–Retest Control the rating positions of all stimuli except FOG, STEEL, FAM, and CEF were randomized.

TABLE 1

SUMMARY OF DESIGN AND PROCEDURE

Experimental Groups	Time 1 Rating	Relations Training	Response Training	Time 2 Rating	Analysis Groups: Amount of Change in S_7
Serial	Rate 11 stimuli	Serial learning 12 syllables	VED (S_7) associated with shock	Rate 11 stimuli	Maximum change Moderate change Minimum change
Paired-Associates Control	Rate 11 stimuli	Each of 12 syllables learned as responses to numbers	VED associated with shock	Rate 11 stimuli	Maximum change Moderate change Minimum change
Test–Retest	Rate 11 stimuli	No training Interpolated tasks		Rate 11 stimuli	Used to determine regression and to correct change scores

The descriptive terms which labeled the semantic differential scales were SMALL–LARGE, BAD–GOOD*, CALM–AGITATED, THICK–THIN, NICE–AWFUL*, FAST–SLOW, DIRTY–CLEAN*, STRONG–WEAK, PLEASANT–UNPLEASANT*, DULL–SHARP, SAD–HAPPY*, AND HOT–COLD. The scales were in the above order on the page. Each continuous line extended the width of a 9-in. sheet of paper and the only subdivisions marked the positions of the descriptive terms (e.g., *Very Good* and *Very Bad*) which were placed 1.5 in. from each edge. Scores were determined by measuring the distance between the unfavorable end of the evaluative scales and the point checked to the nearest $\frac{1}{8}$ in. Scores were summed over the five evaluative scales marked with an asterisk.

The *S*s were introduced to the rating task as a study of the "variety of meanings and connotations which stimuli have for different people," and were urged to work rapidly. Each stimulus was pronounced by *S* as it appeared in the aperture of a manually operated exposure apparatus. When ratings were completed, *S*s were told that they would be asked to rate another group of stimuli later.

Relations training: Serial condition.—This task was introduced as a test of speed and accuracy of learning and *S*s were instructed in the use of the anticipation method. The *S*s learned a practice list of six nonsense syllables to a criterion of three successive correct trials and then an experimental list of 12 nonsense syllables to the same criterion. A unit was presented on a memory drum every 1.5 sec. with an intertrial interval of approximately 6 sec. The serial position of untrained stimuli-to-be-tested for generalized negative affect was varied systematically among Positions 4 through 10. VED was always in Position 7. When the task was completed *S*s were told they would return to the list later to see how well they could remember what they had learned.

Relations training: PA Control.—The *S*s in this group learned a practice list of six number–nonsense syllable pairs to three successive correct trials and then a list of 12 number–syllable pairs which were: 23-WIX, 29-BIP, 37-YOF, 43-DUZ, 51-NOL, 59-ZUT, 67-VED, 71-DEH, 79-PAF, 83-GAN, and 91-KAQ. Each number appeared for 2 sec., followed by the paired number and syllable for 1 sec. The syllables were the same as those in the serial lists and each pair appeared an equal number of times in random order.

The number of trials given to the pairs in the main experiment for each group of six PA Control *S*s was equal to the median number of trials to criterion for the preceding group of six serial *S*s. This criterion was used in order to insure that the test syllables would be equally familiar to Serial and PA Controls as a result of approximately the same number of exposures to the syllables. These Serial and PA Control groups of six *S*s were also matched on the order in which the syllables were rated at Time 1 and Time 2.

Test–Retest Control.—The *S*s in this group received neither relations nor response training. After completing the first rating of syllables, another *E* administered the following tasks which required about an hour and a half, approximately the same amount of time devoted to relations and response training in the other groups: two parts of the Cooperative School and College Ability Tests of the Education Testing Service; the 100 stimulus words of the Kent-Rosanoff list to which *S*s responded with free associations; a task requiring judgments of difference or similarity between pairs of meaningful words and pairs

of shapes; and a test involving the production of synonyms and antonyms of a list of words. After the tasks were completed, *S*s made a second rating of the words and nonsense syllables (see below).

Response training.—A single syllable (always VED) which had been included in relations training, and other nonlist stimuli were associated with shock in both the Serial and PA Control conditions

Two telegraph keys were placed in front of *S* who was told to hold down the left key until a "word" was presented in the frame. When the "word" appeared she could either (*a*) continue to press the left key or (*b*) release the key and press the right-hand key (using one hand). She was told that she would be paid for each correct response and that she might accumulate between $1.50 and $2.00 which she could keep.

The *E* then explained that two kinds of stimuli would be presented. For *punishment* stimuli *S* received a shock regardless of what she did, but by pressing the right-hand key she could turn off the shock. However, in order to be paid on that trial she had to release the left key *before* the shock was turned on, approximately 1 sec. after the presentation of the "word"; the shock continued until *S* pressed the right-hand key. For *nonpunishment* stimuli *S* never received a shock but was paid only if she did *not* release the left key.

Before training began the intensity of shock was adjusted to a level reported to be "painful." The *S* was told that at any time she could request a reduction in the intensity of shock, but that the purposes of the experiment required that it be "a punishment." During training, shock intensity was increased until *E* judged from *S*'s reactions that the shock was painful. At or before this point *S*s often requested a decrease in shock, from which decreased level the program of gradual increases was resumed. Recognition of change in intensity was explained to *S* as resulting from "increased sensitivity." If she refused to be shocked or became upset the experiment was terminated. The shock was administered through a cuff fastened to *S*'s nonpreferred arm and was generated by a Phipps-Bird inductorium or by a Harvard Apparatus Company inductorium, No. 9358.

All *S*s included in the results received at least 84 response training trials, 42 of which were punishment trials. The order of syllables was random within each set of five or six trials. The syllable VED, always associated with shock, and CEF, never associated with shock, both occurred once within each set of trials. Fourteen other syllables were presented, some followed by shock and some not. In addition to VED the following syllables were associated with shock: MEC, GUH, GIH, MUH, HES, MIG, NAH. In addition to CEF, the following were not related to shock during response training: LIH, HUC, HEG, NIH, CUH, NEH, GAH. If *S* responded correctly to both VED and CEF on five successive sets of trials, training was stopped. If *S* had not met this criterion within the first 84 trials the last four sets of trials were repeated until either (*a*) the above criterion was met; (*b*) VED and CEF were responded to correctly on four successive sets beginning with Trial 60, 85, or 109; or (*c*) 133 trials had been completed, whichever occurred first.

Affective rating: Time 2.—In all conditions the same nonsense syllables which had been rated at Time 1 were presented in the same order, preceded by two different meaningful words. The *S*s were told before the ratings that some words might be the same as those rated before but that most would be different. In Serial and PA Control conditions the shock cuff was left on *S*'s arm during these ratings and the inductorium continued to hum, but Time 1 and Time 2 rating conditions were the same in other respects.

Reliability of ratings was determined by data from the test–retest group. When each *S*'s rating of each of the seven nonsense syllables was treated as an independent case, the correlation between Time 1 and Time 2 ratings was .448. Since for 30 *S*s there was a correlation of .516 between mean Time 1 rating of seven syllables and mean Time 2 rating, this estimate of individual syllable rating reliability was based to some extent on between-individual rather than between-stimulus variance. Therefore, in order to estimate individual syllable reliability, scores were normalized for each *S* within each time of rating, and the resulting correlation between Time 1 and Time 2 *z* scores was .407. These reliabilities are lower than those usually reported for semantic differential ratings of meaningful stimuli. Both the use of nonsense material and of a continuous scale probably contributed to the lower reliability. Reliability was intentionally sacrificed since it was expected that memory of the ratings would be reduced with an unmarked scale. It was hoped that these procedures would result in increased sensitivity of the measure.

Definition of trained change groups.—Since it was predicted that generalization would be a function of amount of change in S_7, Serial and PA Control *S*s were sorted in the follow-

ing manner: (*a*) those who did not change their rating of the trained stimulus or who changed it positively were not used in the major analyses; (*b*) three groups of approximately equal size and differing in the amount of corrected negative trained change constituted the sample used unless otherwise indicated. These were designated as the ***maximum trained change***, ***moderate trained change***, and the ***minimum trained change*** groups (see Table 1).

Results

Before proceeding to the tests of hypotheses two questions will be considered:

1. How well were interfering variables equated?

2. Was the necessary condition for a test of generalization met; that is, did the rating of the trained stimulus become more negative following its association with shock? After these questions are answered, the hypotheses will be tested by three types of analysis: (*a*) a comparison of change in affective ratings of stimuli three positions backward and forward from the trained stimulus in the Serial condition with mean change in ratings of untrained stimuli in the PA Control condition; (*b*) a comparison of change in ratings of test stimuli in each serial position with all other positions within the Serial condition; and (*c*) a comparison of change in ratings of each test stimulus in a given serial position with a stimulus in a corresponding position at other levels of trained change.

Equating Variables

In order to increase the confidence with which results can be attributed to the experimental variables (experimental condition, serial position, and amount of change in the trained

TABLE 2

Equation of Variables

Variable	Difference between Serial and PA Conditions at Each Serial Position within Trained Change Groups			Difference between Trained Change Groups at Corresponding Serial Positions			Difference between Serial Positions within Serial Condition and Trained Change Group		
	Max.	Mod.	Min.	Max.-Mod.	Max.-Min.	Mod.-Min.	Max.	Mod.	Min.
Mean Time 1 rating of untrained stimuli at each serial position	ns	ns	ns	ns	ns	ns	S_4-S_7 >0 $P<.05$	S_5-S_7 >0 $P<.05$	ns
Mean Time 1 rating of trained stimulus	ns	ns	ns	>0 $P<.05$	ns	ns	=[a]	=	=
Mean corrected change in trained stimulus	ns	ns	ns	Significant by selection			=	=	=
No. relation training trials	ns	ns	ns	ns	ns	ns	=	=	=
No. response training trials	ns	ns	ns	ns	ns	ns	=	=	=
Order of rating	See text			See text			Not tested		

[a] "=" means equated by experimental design.

stimulus), it is necessary to determine whether certain other variables were equated. Table 2 indicates five variables which were tested in order to determine whether they had significantly different values in the groups compared in the major analyses.

Time 1 ratings of stimuli.—The principal dependent variable of the experiment was the change in attitude scores determined by subtracting the Time 1 evaluative rating of each nonsense syllable from the Time 2 rating. It was immediately obvious that the uncorrected change score was influenced by the initial rating of a nonsense syllable. Those syllables which were initially positive were more likely to change in a negative direction, while the initially negative tended to become more positive at Time 2. The magnitude of this regression factor was estimated from the Test–Retest Control data which showed a correlation of −.507 between initial ratings and uncorrected change scores. Using a regression equation based on this correlation, all initial ratings of each *S* were converted to expected change scores. The result of subtracting each expected change score from the uncorrected change score was designated the *corrected change score*, with a negative score meaning that *S* changed her rating of the syllable more unfavorably than would have been predicted from knowing her initial rating. All change scores have been reported in this form unless otherwise indicated.

But differences in Time 1 scores might have other effects on attitude change which were not eliminated by the correction described. It is of importance, therefore, to examine the extent to which Time 1 ratings were equated. Row 1 of Table 2 shows that only two of the comparisons for untrained stimuli were significant; namely, the comparison of S_4 and S_7 in maximum change group and S_8 and S_7 in moderate change group. In Row 2 (trained stimuli) one was significant; the Time 1 rating of S_7 in the maximum change group was less positive than the Time 1 rating of S_7 in the moderate change group. One set of comparisons omitted from the Table also revealed a significant Time 1 difference: that between the rating of S_7 maximum and S_7 moderate under PA Control conditions. In other words, only four out of 108 comparisons involving Time 1 ratings of syllables in different serial positions, in different experimental conditions and at different levels of trained change were significant. Comment will be made in the discussion section on the possible relation of one of these significant differences to certain results.

Number of relations training trials.—For comparisons between serial positions within a trained change group the number of trials was necessarily the same. There were no significant differences between Serial and PA Control conditions within any level of trained change or between levels of trained change (see Row 4, Table 2).

Number of response training trials.—Row 5 of Table 2 shows that this variable was also equated in all the comparisons made.

Average rating order of stimuli.—Since the order of rating stimuli may influence judgments, it is important that this factor be equated for stimuli in different serial positions, in groups with different amounts of trained change and in Serial and PA Control groups. Because the number of cases within each cell was not large enough to make a satisfactory test of the difference in the distribution of rating positions for each condition, it has been necessary to determine whether the findings were changed when rating order was held constant. These analyses are presented later.

Change in Response to the Trained Stimulus

The total sample was used to test whether ratings of S_7, the stimulus associated with shock, became more negative. In both Serial and PA Control conditions the mean corrected change score of the trained stimulus was significantly more negative than was the mean corrected change score for the untrained stimuli. In the Serial condition all six comparisons had Ps $< .05$. In the PA Control condition the difference between the mean change in S_7 and the mean of individual average change scores of the six untrained stimuli was significant at the .001 level.

As already explained, it was expected that the amount of generalized attitude change would be a function of the amount of change in S_7 and the effects of serial position have been analyzed with change in S_7 held con-

TABLE 3

GENERALIZED CHANGE IN AFFECTIVE RESPONSES

Condition	Corrected Change Score for Trained Syllable											
	Minimum				Moderate				Maximum			
	N	Mean	*SD*	Diff.[a]	*N*	Mean	*SD*	Diff.[a]	*N*	Mean	*SD*	Diff.[a]
Serial												
S_4	16	0.9	13.0	−1.9	18	3.4	12.8	− 1.4	15	3.6	19.3	0.4
S_5	16	2.0	12.8	−0.8	18	6.2	15.7	1.4	15	3.4	17.8	0.2
S_6	16	2.0	11.5	−0.8	18	− 1.0	12.1	− 5.8*	15	17.0	11.9	13.8***
S_7 (Trained)	16	−7.0	2.5	−1.0	18	−14.0	1.8	0.0	15	−30.0	8.1	−5.0
S_8	16	5.8	13.6	3.0	18	− 6.8	16.4	−11.6**	14	− 5.0	14.6	−8.2*
S_9	16	0.4	12.1	−2.4	18	7.8	12.3	3.0	15	2.9	16.4	−0.3
S_{10}	16	4.6	12.1	1.8	18	9.0	13.2	4.2	15	0.6	15.9	−2.6
PA Control												
Untrained	11	2.8	4.1		13	4.8	4.9		15	3.2	7.1	
Trained	11	−6.0	2.6		13	−14.0	2.4		15	−25.0	9.6	

[a] Algebraic difference between Serial and PA Control; Untrained Serial compared with Untrained PA Control average, and Trained Serial compared with Trained PA Control.

* $P < .10$.
** $P < .02$.
*** $P < .01$.

stant at different levels (see Table 1). At none of the three levels of change in S_7 was there a significant difference between Serial and PA Control in the amount of change in the trained stimulus (Table 2, Row 3).

Generalization as a Function of Position within a Serial Structure

Change in attitude toward the untrained stimuli which were serially related to S_7 has been compared to the mean change of untrained stimuli in the PA Control group with level of trained change held constant. This comparison has also been made holding rating position constant at each level of trained change. Finally, change at each serial position has been compared with change at every other position, again with trained change held constant.

Comparison of change under Serial and PA Control conditions.—Hypotheses 2a, 3a, and 4a predicted that change in S_6 and S_8 would differ from change in untrained PA Control stimuli. Mean corrected change scores for Serial Positions S_4 through S_{10} in each of the three trained change groups are presented in Table 3 and Fig. 6. Entries for the untrained PA Control stimuli are means of individual average change scores for the six test syllables. The significance of differences between means for each list position and the mean of the untrained Control stimuli was tested at each level of trained change.[5]

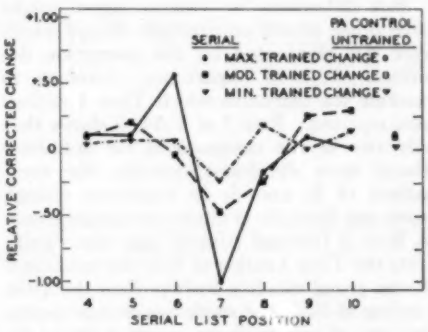

FIG. 6. Change in affective rating as a function of serial position. The ordinate is corrected change, divided by the absolute value of corrected change for S_7 in the maximum trained change group.

[5] Where estimated variances were heterogeneous, the approximation to *t* described by

For *Ss* with maximum change in S_7 the rating of S_6 changed more positively than the average of PA Control ratings of all untrained test stimuli ($P < .01$)[6] and S_8 changed more negatively ($P < .10$). At the level of moderate trained change, ratings of S_6 tended to become more negative than PA Controls ($P < .10$), while S_8 scores were significantly more negative than PA Control scores ($P < .02$). There was no significant difference between any other serial Position and PA Control change score.

The results just reported have tested the hypotheses in question by comparing the amount of corrected change in attitudes from Time 1 to Time 2. However, a part of the theory from which predictions were derived involves assumptions about events at the time of the second rating of the stimuli. This means that the relations must have existed at Time 2. Since Time 1 ratings for the various stimuli and conditions were significantly different in only four of the 108 comparisons, it is reasonable to assume that differences in change scores will reflect differences in the Time 2 ratings of different stimuli. Nevertheless, this has been checked by computing mean Time 2 ratings for S_6, S_7, S_8 within each trained-change level and for the PA Controls trained and untrained within each level. The results obtained by this analysis were exactly parallel to those reported in terms of amount of trained change. The same comparisons were significant and the same ones fell below acceptable levels of significance.

Before stating the conclusions of these tests of the Hypotheses we turn, then, to one other check; namely, the effects of holding constant the order in which stimuli were rated.

Comparison of change under Serial and PA Control conditions with rating order constant.—It will be recalled that it was considered necessary to check the effects of our principal variables with order of rating the test stimuli held constant because of the impossibility of testing adequately whether this variable was equated. Table 4 presents results. They were essentially the same as those just reported.

This table was constructed in the following manner. Within a given level of trained change, mean change in S_6 was computed for Serial *Ss* who rated this stimulus at each possible rating position (4, 6, 7, 8, 9, or 10). The average of PA Control test stimuli rated at each of these positions was also computed for the comparable trained change group. Using these values from Table 4, the PA Control mean at each rating position was subtracted from the Serial mean for S_6 in that same rating position. The same procedure was followed for S_8 change scores within each level

Dixon and Massey (1951, pp. 104–105) was used in testing significance. Since the measure of change for test syllables in the Controls was an average of six scores, its variance was lower than variance of the measures for the Serial condition. A two-tailed test was used throughout. A number of additional predicted relations would be significant at the .05 level if a one-tailed test were used.

[6] These results were based on all *Ss* whose Time 2 ratings changed negatively, even though the hypotheses involving S_6 were not made explicit until one-third of the *Ss* had been run. It should be noted, however, that S_6 results based *only* on the *Ss* who were run following the revision of Hypotheses 3 and 4 were exactly parallel to those reported in the tables. The change in S_6 for the maximum change group was significantly ($P < .02$) more positive than the mean change of untrained stimuli in the PA Controls, maximum change group. Change in S_6 in moderate change group was more negative than for PA Control moderate change group ($P < .10$). For the minimum change group, $P > .10$.

of trained change. The significance of each set of six differences (where there were no empty cells) has been evaluated using a randomization test for matched pairs (Siegel, 1956, pp. 88–92).

With rating order held constant, the relations between corrected change scores of stimuli at different serial positions were maintained with surprising consistency considering the very small number of cases in each cell. As indicated in the Table, mean differences between Serial and PA Control conditions were significant for comparisons of S_6 and S_8 in the maximum and moderate change groups, respectively, but not significant ($P < .10$) for the other two comparisons.

The results strongly support Hypothesis 3a, while Hypotheses 2a and 4a received moderate support.

Comparison of change at different serial positions.—Hypotheses 2b, 3b, and 4b predict different amounts of generalization *within* a serial list. For each *S* in the Serial condition the difference between corrected change scores for each pair of list positions was computed and averaged over *S*s within each level of trained change.

The results in Table 5 indicate that when S_7 changed maximally, S_6 showed significantly more positive change than syllables at every other position, while S_8 was consistently more negative than all others except S_7 but significantly different only from S_6 and S_7.

At the level of moderate trained change S_6 changed more negatively than S_4, S_5, S_9, and S_{10}, the last three differences having *P*s of .10 or less; S_8 was more negative than every other test stimulus at this level and only in the comparison with S_6 and S_7 did the difference fail to reach the .05 level of significance. None of the comparisons involving pairs which did not include S_6, S_7, or S_8 were significant and no differences were significant in the minimum trained change group except those involving S_7.

TABLE 4

MEAN CORRECTED CHANGE FOR TEST SYLLABLES AT EACH RATING POSITION IN SERIAL AND PA CONTROL CONDITIONS

Trained Change Group	Test Syllable	Rating Position						Serial − PA Control	
		4	6	7	8	9	10	Mean Diff.	*P**
Maximum	Serial S_6	31.0	9.0	20.0	20.0	12.0	22.0	15.7	<.04
	PA Control	− 5.2	4.4	10.3	1.0	3.5	5.5		
	Serial S_8	6.7	−11.0	−11.5	−14.0	−14.0	2.7	−10.1	<.10
Moderate	Serial S_6	− 1.5	− 7.5	10.3	− 9.0	7.0	− 6.0	− 5.9	<.10
	PA Control	− 0.9	6.2	6.3	4.1	12.5	0.4		
	Serial S_8	− 4.0	−13.8	− 0.7	− 4.7	− 9.0	− 1.0	−10.2	<.04
Minimum	Serial S_6	1.5	− 1.0	*	2.2	10.0	1.7	− 0.3	ns
	PA Control	4.4	− 0.7	1.3	7.1	− 0.2	5.1		
	Serial S_8	7.6	*	27.5	− 7.5	0.5	2.8	2.6	ns

* By a randomization test on differences at rating positions.
* No cases in the Serial condition. This rating position was omitted in tests of significance.

TABLE 5

MEAN DIFFERENCE IN CORRECTED ATTITUDE CHANGE FOR PAIRS OF SERIAL POSITIONS

Serial Positions	Trained Change Group								
	Maximum			Moderate			Minimum		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
S_7-S_4	15	−33.3****	23.3	18	−17.4****	12.4	16	− 7.6**	12.0
S_7-S_5	15	−33.1****	18.2	18	−20.2****	15.0	16	− 8.8**	12.2
S_7-S_6	15	−46.7****	15.0	18	−13.0****	12.2	16	− 8.8**	11.7
S_7-S_8	14	−24.1****	11.2	18	− 7.2	17.4	16	−12.6****	14.3
S_7-S_9	15	−32.6****	18.1	18	−21.8****	12.1	16	− 7.2**	11.8
S_7-S_{10}	15	−30.3****	15.8	18	−23.1****	13.5	16	−11.3****	12.4
S_6-S_4	15	13.4**	23.2	18	− 4.4	14.7	16	1.1	14.2
S_6-S_5	15	13.6**	21.2	18	− 7.2*	15.6	16	0.0	16.2
S_6-S_8	14	20.7****	17.3	18	5.8	17.4	16	− 3.8	21.2
S_6-S_9	15	14.1****	17.3	18	− 8.8*	17.6	16	1.6	11.2
S_6-S_{10}	15	16.4****	19.5	18	−10.0**	18.7	16	− 2.6	13.8
S_8-S_4	14	− 7.8	25.4	18	−10.2****	14.9	16	4.9	20.7
S_8-S_5	14	− 8.4	21.2	18	−13.0***	24.3	16	3.8	22.7
S_8-S_9	14	− 5.9	17.2	18	−14.6***	23.6	16	5.4	20.1
S_8-S_{10}	14	− 5.4	17.6	18	−15.8****	20.9	16	1.2	18.4
S_4-S_5	15	0.2	32.7	18	− 2.8	19.4	16	− 1.1	13.2
S_4-S_9	15	0.7	21.3	18	− 4.4	17.2	16	0.4	14.7
S_4-S_{10}	15	3.0	17.4	18	− 5.6	17.8	16	− 3.7	18.1
S_5-S_9	15	0.5	27.9	18	− 1.6	20.1	16	1.6	15.3
S_5-S_{10}	15	2.8	27.7	18	− 2.8	23.8	16	− 2.6	19.3
S_9-S_{10}	15	2.3	17.6	18	− 1.2	20.8	16	− 4.1	12.0

* $P < .10$.
** $P < .05$.
*** $P < .02$.
**** $P < .01$.

The results parallel those of the previous analysis: Hypothesis 3b was strongly supported while 2b and 4b received support at a borderline level. In all analyses, S_8 in the moderate trained change group was significantly negative as had been predicted for S_8 in the maximum trained change group, while in the latter group there was less significant negative change.

Generalization as a Function of Change in the Trained Stimulus

Hypotheses 1 and 3c predict that amount of generalization within a serial structure is a function of level of change in response to the trained stimulus. A direct comparison of Serial and PA conditions at different levels of trained change could not be made because of inhomogeneity of variance. In the PA condition, however, differences between change levels in average corrected change scores for the untrained stimuli were small and not significant.

Interlevel comparisons of change in S_6 and S_8.—The comparison of mean corrected change for corresponding serial positions at the three levels of trained change (Table 3) indicated that change in S_6 was significantly more positive when S_7 changed maximally than in either

the minimum or moderate change conditions ($P < .005$ in both cases). Change in S_6 was not significantly different between the latter two levels.

The same comparisons for S_8 showed significantly more negative change in the moderate than the minimum condition ($P < .05$) but the predicted difference in the same direction between the maximum and minimum change groups did not reach an acceptable level of significance with a two-tailed test ($P < .10$). The difference between the maximum and moderate change groups was opposite that predicted and was not significant.

Interlevel comparisons of change in S_6 and S_8 with rating order held constant.—Differences between levels of trained change in mean S_6 scores at each rating position (computed from Table 4) were evaluated using the randomization test for matched pairs (Siegel, 1956, pp. 88–92). The same procedure was followed for S_8. By this analysis all differences in S_6 change between levels of trained change were significant or near significant (maximum > moderate, $P > .04$; maximum > minimum, $P < .07$; and moderate < minimum, $P < .07$). Only one difference for S_8 approached significance (maximum < minimum, $P < .07$). The symbol (<) means more negative; (>) more positive. These results differ somewhat from those where rating order was not controlled. The near significant difference for S_6 between moderate and minimum change groups was not observed there, and this is the only analysis where the significance of change for S_8 was greater in the maximum than in the moderate trained change groups.

Interlevel comparison of difference between serial positions in attitude change.—Essentially the same results were obtained when difference between change in ratings of S_6, S_8, and other untrained list syllables was investigated as a function of change in S_7 (Table 5). The mean difference between S_6 and each other serial position at the maximum level of change in S_7 was subtracted from the mean difference between S_6 and that same position at a different level of trained change, and the significance of the difference determined. This comparison was made between levels for every difference involving S_6 and S_8.

All five differences involving S_6 were significant ($P < .05$) in comparisons of maximum and moderate levels of trained change; in maximum–minimum comparisons three were significant at the .05 level and one at the .10 level; one was significant at the .10 level in comparison of moderate and minimum trained change groups.

For S_8 only comparisons between moderate and minimum levels of trained change were significant (four out of five at the .05 level). In all comparisons of the maximum and minimum groups the differences were in the expected direction: S_8 was relatively more negative than other positions in the maximum than in the minimum group. Here, as in all but one other analysis, the greater amount of negative change in the moderate than in the maximum change group was apparent, though the differences were not significant.

It may be concluded from all methods of analysis that there was strong support for Hypothesis 3c in the greater positive change in S_6 when S_7 changed maximally. Hypothesis 1 was supported by the fact that greater negative change in S_8 occurred when change in S_7 was moderate than when it was minimum. However, the general finding of slightly more nega-

tive change of S_8 in the moderate than in the maximum trained group was not predicted.

Discussion

The hypotheses have been supported in general except for 1 and 2a, which predicted the greatest negative change in S_8 when S_7 changed maximally. Actually S_8 changed more negatively in the moderate group than in the maximum change group. Two possible reasons for this result suggest themselves. First, there is a possibility of the operation of an artifact incident to the significantly more positive Time 1 rating of S_7 in the maximum change group than in the moderate group in both Serial and PA Control conditions. The question is whether this different Time 1 rating has made the amount of trained change an appropriate basis for choosing groups since under these circumstances the Time 2 rating of S_7 in the moderate change group *might have been* more negative than the Time 2 rating of the maximum change group. If this were so, then it would be expected that the Time 2 rating of S_8 moderate might be more negative than the rating of S_8 maximum. A check on the mean Time 2 ratings of S_7 in each group indicates that this was not the case. The Time 2 ratings in the maximum, moderate, and minimum trained change groups had the same order indicated by amount of trained change; moreover, the differences in Time 2 scores for the three groups were highly significant, all with probabilities less than .001.

Still another hypothesis is plausible. It will be recalled that our prediction about the effects of events occurring at the time of testing took into account the possible influence only of the immediate forward associate of the stimulus being tested. This was recognized as an oversimplification which, as it now appears, may need reconsideration. It may be that *multiple* events elicited by a stimulus combine in various ways to influence the rating which *S* gives to a stimulus. Only two possible loci of such a combination can be mentioned at this time, both of which should have similar effects. First, assume that *either* the immediate forward or the immediate backward associate is elicited on any one occasion with the probabilities suggested by the Raskin and Cook data. Then the forward associate will provide the context of rating a given stimulus on some proportion of the semantic differential scales and the backward associate will be the context for the ratings on other scales (five evaluative scales were checked by each *S* for each stimulus). Or, secondly, it may be that *each* judgment on a single scale is an average or some combination of interactions between $r_{s,n}$ and $r_{s,n+1}$ and interactions between $r_{s,n}$ and $r_{s,n-1}$ all of which are involved in the process of judging the position of S_n on one scale. Again, the frequencies of each of these two types of interaction would be proportional to the probabilities of the elicitation of a forward and a backward associate, respectively.

Take as an example the situation under discussion, the rating of S_8 in the moderate and the maximum change groups. Suppose that S_8 when presented for rating elicits $r_{s,8}$ which in turn elicits $r_{s,9}$ and/or $r_{s,7}$ in the ratio of about 30/7 which corresponds to the relative frequency of occurrence of immediate forward and backward associates. The suggestion is that the judgment of S_8 would be weighted more heavily by the results of interaction of $r_{s,8}$ with $r_{s,9}$ than by the interaction of $r_{s,8}$ with $r_{s,7}$ (whether the average is over scales or over events involved in a single judgment or over individuals). Now, since the $r_{s,7} - r_{s,8}$ distance would more likely reach contrast levels when S_7 change is maximal, this would be the condition under which the judgment of S_8 would be most influenced in the *positive* direction by the elicitation of $r_{s,7}$. On the other hand, at moderate change levels elicitation of $r_{s,7}$ would more probably lead to assimilation of $r_{s,8}$ to $r_{s,7}$ which would tend to make S_8 more *negative*. In short, under conditions of maximum change in S_7 the contribution of the backward

associate to the judgment of S_s would be to make it more positive and under conditions of moderate trained change the contribution of the elicited backward associate would be to make the judgment of S_s more negative.

This *post hoc* suggestion about the possible effects of backward associates occurring at the time of testing for generalization serves to introduce a final comment about the complexities of the problem of generalization. We have considered *one* structural pattern (serial associates), *one* aspect of structural change taking place as a result of response training, and the interaction of *one* set of events at the time of testing: the affective response associated with the test stimulus at the time of response training, the forward associate of the test stimulus and its attached affect. It is at once evident that change in conditions at each of these three points will alter the response made to any stimulus tested for generalization. First, there are other structures within which generalization occurs and these will differ in the number of related parts and in the probabilities of arousal of each part by every other part.

Secondly, at the time of response training the structure may be changed in a variety of ways by different input patterns operating on initially different structures. To take but one example: we have some evidence (preliminary and unpublished) that when a positive affective response is attached to one point in a serial structure and negative affect to another point in the same structure, the serial relation no longer plays the dominant role in determining the occurrence of generalization from one point to another within the serial structure. Finally, the response elicited at the time of testing for generalization may be altered by the interaction of a variety of concurrent events initiated by the particular instructions, by the pattern of contextual stimuli, and by the whole matrix of associations that may be activated by the stimulus-to-be-tested as a result of existing structural arrangements.

A concluding comment must be made regarding the relation of the theory here presented to the more conventional mediation theory. Although the results reported at this time are not those that would have been predicted by the mediation theorists, there is no doubt that under some conditions presentation of an untrained stimulus will have some probability of being followed by the response which has been associated with the forward associate of the test stimulus and this may occur because at the time of testing the untrained stimulus elicits the "mediating" forward associate which in turn elicits the response. But neither is there any doubt that generalization may appear under other conditions. It is our suggestion that a complete theory of generalization must specify the different conditions under which generalization occurs. And this may be accomplished only by taking into account the structural arrangements and the input patterns operating both at the time of training the response and of testing the untrained stimuli for evidences of generalization.

Summary

A theory is presented which attributes the nature of the response made to a stimulus being tested for generalization within a serial structure to the combined effects of (*a*) associations formed at the time that the stimulus-in-training is being associated with a response, and (*b*) the interaction at the time of testing between the response to the test stimulus and responses to other events associated with the test stimulus.

The *S*s rated 12 nonsense syllables on evaluative scales. One group then learned the nonsense syllables in serial order after which the seventh syllable was associated with shock. Syllables were rated a second time. A PA Control group followed the same procedures except that in place of serial training each of the 12 syllables was the response to two-place number in paired-associates training. An analysis was made of changes in the evaluative ratings of these stimuli for *S*s in three groups: those for whom the change in the trained stimulus was maximal; a moderate trained change group; and a minimum trained change group.

The following results were obtained:

1. Trained stimuli changed significantly in both the Serial and PA Control conditions when testing each group as a whole.

2. The evaluative ratings of S_8 immediately following the trained stimulus changed significantly more in the negative direction than did comparable PA Controls and more than serial stimuli in other list positions, when S_7 changed moderately. The negative change in S_8 in the maximum change group was of smaller magnitude than in the moderate trained change group. This was contrary to predictions.

3. In the maximum trained change group the stimulus immediately preceding the trained stimulus changed more positively than any other serial stimuli and more positively than comparable PA Controls, as predicted. This positive change in S_6 disappeared as the change in the trained stimulus decreased and tended to be negative at the moderate level of trained change. This was expected.

It is proposed that the positive change in S_6 at maximum trained change levels resulted from contrast with the very negative implicit response to S_7 which was elicited at the time of testing S_6. The negative change in S_8 was predicted on the basis of events assumed to take place at the time of response training.

REFERENCES

BUGELSKI, B. R. A remote association explanation of the relative difficulty of learning nonsense syllables in a serial list. *J. exp. Psychol.*, 1950, **40**, 336–348.

DIXON, W. J., & MASSEY, F. J. *Statistical analysis.* New York: McGraw-Hill, 1951.

MINK, W. D. Semantic generalization as related to word association. Unpublished doctoral dissertation, Univer. of Minnesota, 1957.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *Measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

PEAK, H. Psychological structure and person perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford Univer. Press, 1958. (a)

PEAK, H. Psychological structure and psychological activity. *Psychol. Rev.*, 1958, **65**, 325–347. (b)

RASKIN, E., & COOK, S. W. The strength and direction of associations formed in the learning of nonsense syllables. *J. exp. Psychol.*, 1937, **20**, 381–395.

RAZRAN, G. H. S. Semantic and phonetographic generalizations as a function of salivary conditioning to verbal stimuli. *J. exp. Psychol.*, 1949, **39**, 642–652.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

(Received May 27, 1959)

Journal of Experimental Psychology
Vol. 59, No. 5, 1960

TRANSFER OF A DISCRIMINATION FROM OBJECTS TO PATTERNS [1]

BETTY J. HOUSE AND DAVID ZEAMAN

University of Connecticut

Lawrence (1952) presented evidence for an empirical law of discrimination learning that, in training a hard discrimination, it is more efficient to train first on easier discriminations along the same dimension than to give all trials on the hard discrimination. After equal numbers of trials, rats with all training on a difficult brightness discrimination performed more poorly than those having trials first on easier brightness discriminations. Baker and Osgood (1954) observed similar effects with adult humans in pitch discrimination. These results provide an exception to the usual finding of transfer studies that maximum performance is obtained when training and test conditions are identical. In the present study, an attempt is made to extend the generality of the law by testing it with a different kind of *S*, mentally retarded children, and a different type of problem, color-form discrimination.

A difficult discrimination problem for mentally retarded *S*s is one in which the stimuli are patterns painted on a background. An easier problem can be obtained by presenting stimuli as three-dimensional objects. Even in the case where the relevant cues are identical, an object discrimination is learned more quickly than a pattern discrimination by retarded *S*s (House, Zeaman, Orlando, & Fischer, 1957) as well as by monkeys (Harlow, 1945). Thus, preconditions for a test of the law demonstrated by Lawrence are present: an easy and a hard discrimination problem with the same relevant dimension.

It might be questioned how object and pattern stimuli can be ordered along the same stimulus dimension, as required for application of Lawrence's law. In answer, the two object stimuli may be regarded as differing along a continuum of partial identity. The positive and negative objects are composed of some common stimulus elements as well as distinctive relevant cues. The present pattern stimuli are conceived as having the same relevant elements as the objects, but with an added number of common background elements. Thus the pattern stimuli occupy intermediate positions between the two objects along a continuum of partial identity, conceivably measurable in terms of overlap of sets of elements.

In this experiment, discrimination was trained between stimuli differing in color and form, presented in either of two ways: as flat patterns on gray backgrounds or as cut-out three-dimensional objects. One group (P-1) was given trials on the pattern stimuli only. Another group (0–1 to P–1) was trained first on a pair of object stimuli before being shifted to a pair of pattern stimuli having the same colors and forms as the

[1] This research was carried out at the Mansfield State Training School, Mansfield Depot, Connecticut. The authors wish to express appreciation to Neil A. Dayton, Superintendent, and to John Cassell and Louis Boly, principal and assistant principal of the Longley School, for their aid and cooperation. This investigation was supported by Research Grant M-1099 from the National Institute of Mental Health.

TABLE 1

SUBJECT CHARACTERISTICS OF THE EXPERIMENTAL GROUPS

Condition	*N*	CA (Mo.)		IQ		MA (Mo.)	
		Mean	*SD*	Mean	*SD*	Mean	*SD*
0–1:P–1	19	128	34	39	12	47	12
0–2:P–1	22	136	35	41	11	53	13
P–1	19	132	40	40	13	49	12

objects. This condition is analogous to Lawrence's easy-to-hard group. A third group (0–2 to P–1) was trained on object stimuli different in color and form from the pattern stimuli. In this condition, approach-avoidance tendencies acquired to the object stimuli cannot be directly transferred to the new pattern stimuli.

METHOD

Apparatus

The discrimination apparatus consisted of a table with a one-way screen separating *E* from *S*, and a sliding tray for presentation of the stimuli. Two 3-in. foodcups were set into the 30-in. tray, centered 12 in. apart. The tray was pulled out of sight behind the one-way screen for baiting and pushed diectly in front of *S* to begin a trial.

Two pairs of color-form stimuli were constructed in the form of object stimuli and pattern stimuli. These were (*a*) yellow T vs. black square and (*b*) red cross vs. green circle. For the object stimuli, forms were cut from ¼-in. Masonite and mounted vertically on 4-in. × 4-in. gray Masonite bases. For the pattern stimuli, forms of the exact dimensions and colors of the object stimuli were painted on gray wedges with 4-in. × 4-in. faces. Each form had maximum height and width of 2 in.

Subjects

All *S*s were experimentally naive, selected randomly from boys and girls attending classes at the Longley School of the Mansfield State Training School, within the mental age range from 2 yr. to 6 yr. The *S*s were assigned to the three groups so as to equate for CA, IQ, and MA, as shown in Table 1.

Procedure

General.—The *S*'s response was lifting the stimulus. Under the correct stimulus, *E* placed a small piece of candy. Reinforcements were M & M's, candy corn, Malties, and the like, which *S* could eat or save as he preferred. As additional reinforcement, *E* said "Good" if the response was correct and "No" if it was not.

One of the two pairs of stimuli was assigned randomly to each *S* and the positive cue was designated randomly. The correct stimulus was placed on the left or right according to a Gellerman's series (1933). All *S*s were given 25 trials per day with correction procedure. The criterion for learning was 20 out of 25 trials correct during a single daily session.

Pretraining.—Verbal instructions were limited to asking *S* to find the candy. Before the first experimental session, several trials were given in which *E* placed the candy in an open foodcup and then under a single gray wedge placed randomly left or right.

Experimental conditions.—The sequences of training trials for the three groups were as follows. In Group 0–1:P–1 there was object training to criterion with a maximum of 250 trials, followed by identical pattern training to criterion or 250 trials. In Group 0–2:P–1, there was object training to criterion with a maximum of 250 trials, followed by different pattern training to criterion or 250 trials. In Group P–1 there was pattern training to criterion with a maximum of 500 trials.

RESULTS

Performances of the three groups are shown in Fig. 1. The percentage of *S*s who have met criterion on the *pattern* problem is plotted as a function of total training days. For Group 0–1:P–1 and Group 0–2:P–1, total training days refers to the sum of days spent on the object problem *and* the pattern problem. For example, the percentage on Day 2 for Group 0–1:P–1 signifies that 21% learned 0–1 to criterion on the first day and learned P–1 to criterion on the second day. The superiority of the two object-trained groups over Group P–1 is significant. Chi square tests were made of the frequencies of

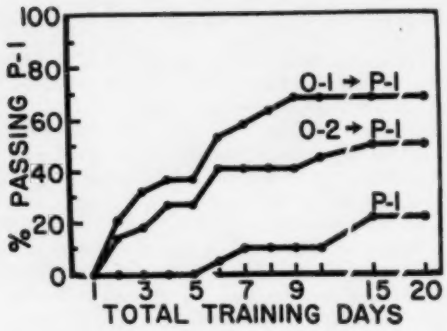

FIG. 1. Percentage of *Ss* in each group who have learned pattern is plotted as a function of days of training. For Group 0–1:P–1 and Group 0–2:P–1, total days refers to sum of object and pattern training days.

Ss who had passed or had not passed the pattern test by the end of 10 days training. The probability that differences as large as those observed would occur by chance is less than .001 for Group 0–1:P–1 and less than .02 for Group 0–2:P–1. Little change occurs during the final 10 days during which all *Ss* who had failed object or pattern were trained on pattern.

While both kinds of object training produced significant facilitation of pattern learning, the question remains whether there is an added advantage when the relevant cues are identical for the object and pattern problems. A chi square test failed to show a difference between percentage passing pattern for the two object groups ($P > .10$). However, when a more sensitive measure was considered, that of number of errors to criterion on the pattern problem, Group 0–1:P–1 was superior ($P < .01$, Mann-Whitney U test, one-tailed). Mean errors were 5 for Group 0–1:P–1 and 47 for Group 0–2:P–1; median errors were 2 and 9, respectively. That the better performance on pattern was not due to chance differences in learning ability is demonstrated by the fact that the two groups were equivalent in performance on the object problem ($P > .70$, Mann-Whitney U test). Percentages of *Ss* learning the object problem were 68 and 63 for 0–1 and 0–2, respectively. All *Ss* in Group 0–1:P–1 who learned the object discrimination also learned the pattern discrimination, while of the 14 *Ss* who learned the 0–2 problem, only 10 then learned the pattern problem.

It is recognized that the curves of Fig. 1 do not reveal the shapes of the underlying learning functions. These are of interest in relation to another theory and will be discussed in detail elsewhere. All the individual functions can be classified as belonging to a family of ogival functions which pass from chance to criterion very quickly, once the curves start to rise. Wide individual differences exist, however, in the duration of the chance stage of performance. Therefore, since *Ss* are for the most part either at chance or criterion, the reported measure of percentage passing is a good approximate description of the state of learning of each group.

DISCUSSION

Our results support Lawrence's conclusion that a training sequence of easy-to-hard discrimination trials with the same relevant dimension is more efficient than a training series of hard discrimination trials only. In addition, our 0–2:P–1 condition demonstrated that direct transfer of approach-avoidance tendencies is not necessary for a facilitating effect. This finding adds weight to Lawrence's (1955) argument that current theories of generalization gradients cannot account for his results. If there is generalization due to the similarity of the 0–2:P–1 stimuli, the effect is as likely to be interference as facilitation since the positive cue is randomly designated. While generali-

zation of approach-avoidance tendencies from the 0–1 training to the P–1 condition presumably accounts for the superiority of 0–1 training over 0–2 training, there remains a large facilitating effect which requires some other explanation.

A model for discrimination learning developed by Restle (1955), which relates learning rate to proportion of relevant and irrelevant cues, has been shown to account for Lawrence's results quite parsimoniously. This theory can be applied to our data to predict the amount of 0–1:P–1 transfer from a knowledge of the learning rate parameters (θ) of the object and pattern learning curves. Parameters were estimated for the four *S*s (i.e., 21%) of Group P–1 who learned the pattern discrimination and for the comparable top 21% (4 *S*s) of Group 0–1:P–1. With mean θs equal to .27 and .02 for the object and pattern *S*s, respectively, Restle's Equation 10 (1955, p. 15) was used to calculate predicted cumulative errors for the four 0–1:P–1 *S*s on the first 25 transfer trials. Table 2 compares the obtained errors with the predicted errors. Although the number of *S*s in this comparison is small, the magnitude of the difference between predicted and obtained errors appears sufficient to warrant rejection of Restle's transfer Equation 10 for these data.

It should be pointed out that Restle's Equation 10 requires the assumption that object and pattern stimuli have the same *relevant* cues, but that pattern stimuli have more *irrelevant* cues than appear in the object stimuli. An alternative assumption could be made: objects have more relevant cues than pattern, but the same irrelevant cues. Equations may then be derived from Restle's theory which predict the almost perfect transfer found in this experiment. Thus, if Restle's theory is applicable in the present experiment, the implication is that in going from 0–1 to P–1 no irrelevant cues are added, but relevant cues are subtracted. The theory makes no predictions about transfer from 0–2 to P–1.

TABLE 2

PREDICTION OF TRANSFER FROM 0–1 TO P–1

Transfer Trials	Cumulative Errors	
	Observed	Predicted
5	2	9
10	2	18
15	2	27
20	2	36
25	3	45

Lawrence (1952) suggested that his results may be due to acquired distinctiveness of the relevant stimulus dimension. Stimuli acquire distinctiveness during easy discrimination training when *S* learns to "isolate functionally the relevant stimulus dimensions from all other background and irrelevant cues" (Lawrence, 1952, p. 516). The distinctiveness then transfers to the difficult discrimination. Details of the process are not elaborated by Lawrence. However, a more detailed analysis of learning of this sort has been developed by Wyckoff (1952), whose model for "observing responses" may account for our results. Wyckoff defines the "observing response" (R_0) as any response which exposes *S* to the relevant cue. For our experiment, we might interpret R_0 as *looking at the color and form cues*. The probability (p_0) of R_0 has been postulated to bear a circular relation to the speed of discrimination learning. Each is dependent upon and facilitates the other. Response R_0 is learned to the extent that its occurrence increases the probability of reinforcement, but in a discrimination problem R_0 can increase probability of reinforcement only if *S* is responding above chance, i.e., is learning.

This reasoning may be applied to our data. For object stimuli, p_0 may be initially higher than for pattern, because of fewer competing background (non-color/form) cues. This condition would facilitate the object discrimination and in turn increase p_0 even further. Consequently *S*s going from object to pattern would have a special advantage in starting pattern with p_0 higher than

that ever developed by Group P–1 during its unsuccessful pattern training.

While Wyckoff's theory may thus be capable of accounting for our nonspecific type of object-to-pattern transfer, his formal model was not designed for the particular situation of the present experiment. A modification of Wyckoff's model is now in preparation which we hope will also handle the quantitative aspects of the data.

Summary

Two different training sequences of easy-to-hard discrimination trials were compared for efficiency with a training sequence of hard discrimination trials only. The *S*s were three independent groups of mentally retarded children equated for MA, IQ, and CA. The hard-discrimination group was trained with pattern stimuli differing in color and form. The two easy-to-hard groups were first trained to criterion with a pair of object stimuli before pattern training. Object training differed for the two groups only in that one group had identical relevant cues for the object and pattern problems while the other group had different colors and forms for object and pattern. Both easy-to-hard sequences were more efficient than the series of hard-discrimination trials only, as measured by total trials required to learn the hard discrimination. The presence of identical relevant cues in object and pattern increased the facilitating effect of object training. Within 10 training days, 68% learned both the object problem and the equivalent pattern problem, while in the other easy-to-hard group, 45% learned an object problem and a different pattern problem. In contrast, only 10% of the group having all training trials with pattern stimuli learned the pattern problem within 10 days. Implications of these data for Restle's discrimination model and Wyckoff's theory of "observing responses" were discussed.

References

Baker, R. A., & Osgood, S. W. Discrimination transfer along a pitch continuum. *J. exp. Psychol.*, 1954, **48**, 241–246.

Gellerman, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *J. genet. Psychol.*, 1933, **42**, 206–208.

Harlow, H. F. Studies in discrimination learning in monkeys: V. Initial performance by experimentally naive monkeys on stimulus-object and pattern discrimination. *J. gen. Psychol.*, 1945, **33**, 3–10.

House, B. J., Zeaman, D., Orlando, R., & Fischer, W. *Learning and transfer in mental defectives. Progress rep. No. 1*, NIMH, USPHS 1957. Res. Grant M-1099 to Univer. of Connecticut. Pp. 58–63.

Lawrence, D. H. The transfer of a discrimination along a continuum. *J. comp. physiol. Psychol.*, 1952, **45**, 511–516.

Lawrence, D. H. The applicability of generalization gradients to the transfer of a discrimination. *J. gen. Psychol.*, 1955, **52**, 37–48.

Restle, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

Wyckoff, L. B., Jr. The role of observing responses in discrimination learning. Part 1. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received May 28, 1959)

Journal of Experimental Psychology
Vol. 59, No. 5, 1960

EYELID CONDITIONING AS A FUNCTION OF INTENSITY OF CONDITIONED AND UNCONDITIONED STIMULI [1]

EVELYN G. WALKER

University of Illinois

Although conditioned stimulus (CS) intensity has theoretically been considered one of the variables determining the strength of conditioned responses (CRs), the experimental evidence is in conflict. While some studies (Barnes, 1956; Brown, 1942; Hovland, 1937; Hull, 1949; Razran, 1957) have shown response strength to increase with increasing intensities of CS, others (Carter, 1941; Grant & Schneider, 1948, 1949; Taylor, 1954) have failed to find any relationship. Those experiments in which response strength was measured during acquisition of the CR typically were successful in demonstrating a positive relationship between CS intensity and strength of response. However, in general, those experiments in which the analysis was based on extinction data have reported negative findings. Within a single experiment (Kessen, 1953), the positive relationship shown for acquisition of a CR could not be demonstrated during extinction. Since the one important difference between acquisition and extinction of a CR is the omission of the unconditioned stimulus (UCS), it may be argued that this factor could, in part, have contributed to the conflicting results. If this is the case, then the effectiveness of CS intensity on response strength may depend upon the strength of the UCS.

One purpose of the present study was to determine whether UCS intensity is one of the conditions which influences the relationship between CS intensity and CR strength. Specifically, the hypothesis was tested that CS intensity would have a greater effect on response strength under a strong UCS than under a weak UCS. Such a result may be deduced from Hull's assumption of a multiplicative relation between habit, drive and stimulus intensity dynamism (V) (Hull, 1951), if UCS magnitude is viewed as a determinant of drive strength. Indirect support for this hypothesis may be found in the work of Pavlov (1927), who found that unless the UCS were strong, ordinarily suitable stimuli were ineffective as CS. In a recent analysis of studies on the conditioned salivary response, Razran (1957) has indicated that unconditioned response (UCR) magnitude is one of the variables related to the efficacy of CS intensity. He reported that the higher the UCR magnitude, the greater the effect of CS intensity on response strength. Since magnitude of the UCR is, in part, a function of UCS intensity, this could be interpreted as indirect evidence that the effect of CS intensity on response strength is related to the level of the UCS. Although in three studies with the eyelid response (Carter, 1941; Grant & Schneider, 1948; Taylor, 1954) no reliable relationship between CS intensity and strength

[1] This paper is based on a portion of a dissertation submitted to the Graduate College of the University of Illinois in partial fulfillment of the requirements for the PhD degree. The writer wishes to express her appreciation to G. Robert Grice, who directed the study, for valuable aid and advice offered throughout the course of the experiment.

of conditioning could be demonstrated, the intensity of the UCS in these studies would appear to have been relatively weak.

If the effect of CS intensity on strength of response is altered by variations in the intensity of the UCS, then it would be difficult to demonstrate such a relationship using the typical extinction procedure where the UCS is omitted. To avoid this problem, the UCS was presented on each trial during extinction in the present experiment but with a CS-UCS interval which has been shown to produce extinction (McAllister, 1953).

A further purpose of the present study was to determine whether CS intensity has an effect on learning as distinct from performance. In his stimulus intensity dynamism principle, Hull (1951) has postulated that both learning (habit strength) and performance (reaction potential) are direct functions of CS intensity. Three experiments (Grant & Schneider, 1948, 1949; Kessen, 1953) have yielded no evidence to support the hypothesis that CS intensity affects habit strength. However, in these studies CS intensity did not influence response strength, and under conditions which fail to show this relationship, there would be little reason to expect an effect of CS intensity on habit strength.

Method

Subjects and design.—The *S*s were 160 men students of introductory psychology at the University of Illinois. Twenty *S*s were assigned at random to each of the eight cells at a $2 \times 2 \times 2$ factorial design. During acquisition, which consisted of 80 paired presentations of tone and air puff separated by 500 msec., there were four different stimulus conditions: weak CS–weak UCS; strong CS–weak UCS; weak CS–strong UCS; strong CS–strong UCS; with two groups under each condition. There were eight conditions during extinction with all groups receiving the same UCS intensity as during acquisition. Four groups received the same CS intensity as acquisition. Two groups were shifted from weak to strong CS intensity, and two groups were shifted from strong to weak CS intensity. For extinction all groups received 30 trials with an interstimulus interval of 2500 msec., the air puff being presented on all trials.

Apparatus.—The CS was a tone of 1000 cycles, with the weak intensity 30 db, and the strong intensity 80 db above an average threshold. The tone was generated by a Hewlett-Packard, Model 205 AG audio oscillator and was delivered to *S* through a set of headphones.

Two intensities of air puff were employed as UCS; 0.5 lb./sq. in. in the case of the weak puff, and 5.0 lb./sq. in. in the case of the strong puff. Appropriate pressures were obtained by the use of a variable air pressure reducing valve. The CS had a duration of 500 msec. during acquisition and 2500 msec. during extinction, with the UCS onset coinciding with the termination of the CS and lasting 40 msec. A polygraph recorded the onset and offset of the stimuli which were controlled by means of Hunter electronic timers. The potentiometer-polygraph system used for recording eyelid movement has been described in detail by Spence (1953).

Conditioning procedure.—The *S*s were seated in a soundproof room. Incident illumination measured at the position of *S*'s eyes was approximately 2.5 ft-c. The ready signal was a weak buzzer which was presented at random intervals of 2, 3, or 4 sec. before each trial. At the ready signal, *S* was instructed to fixate a white disc (located 60 in. from *S* and 3 in. above eye level) and was told to blink once. An average intertrial interval of 20 sec. was used, with trials randomly separated by 15, 20, or 25 sec.

Measurement of CR.—Any response which occurred within 150 msec. of the CS onset was not counted as a CR since this is the interval in which original or unconditioned responses to the onset of the tone typically occur. A movement of the penwriter 1 mm. or more in amplitude in the interval 150 to 565 msec. after the onset of the CS was considered to be a CR. Although the UCS onset was 500 msec. after the onset of the tone, responses which occurred up to 565 msec. after CS onset were included as CRs when it was empirically determined that the minimum latency of the UCR was 75 msec.

In a previous study (Spence, 1953)

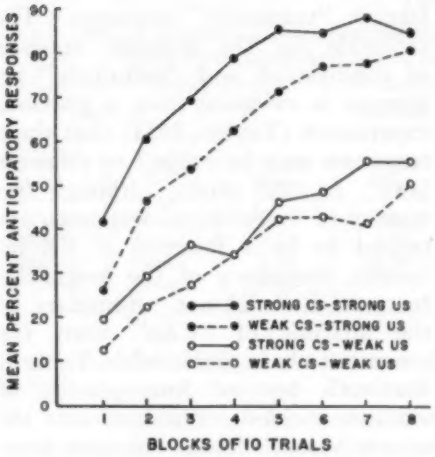

FIG. 1. Mean percentage of anticipatory responses in the interval 150 to 565 msec. during acquisition per block of 10 trials for the four stimulus conditions.

anticipatory responses were differentiated in terms of their latency, with the shorter latency responses being regarded as "voluntary" responses. Under the conditions of the present experiment, responses which occurred in the interval 150 to 300 msec. after CS onset were regarded as short latency or "voluntary" responses. Any *S* for whom more than 50% of his responses were in this interval was discarded as a voluntary responder. Sixteen *S*s were eliminated for this reason and one additional *S* was discarded because of failure to condition, the conditioning criterion being at least one CR during the 80 trials.

During extinction, CRs were also counted in the interval from 150 to 2500 msec. which allowed the inclusion of CRs whose latency had increased beyond the shorter interval (McAllister, 1953).

RESULTS

Acquisition of anticipatory response. —Figure 1 shows the mean percentage of anticipatory responses plotted as a function of blocks of 10 trials. For this analysis, an anticipatory response was defined as one whose latency was in the interval from 150 to 565 msec. after CS onset. Ordering of the groups is clear, the strong UCS groups showing consistently superior performance to the weak UCS groups, and the strong CS intensity groups exceeding the weak CS intensity groups.

A factorial analysis of variance was performed in order to evaluate the reliability of the observed differences in rate of acquisition of anticipatory responses under differential conditioned and unconditioned stimulus intensity. Table 1 summarizes the results of the analysis. The assumption of homogeneity of variance appeared tenable since Bartlett's test resulted in a chi square value which was not significant. It can be seen from Table 1 that the analysis yielded *F* values significant beyond the .01 level for both CS and UCS intensity, with no significant interaction. The test for interaction is a test of the hypothesis that CS intensity has the same effect on performance whether the UCS be strong or weak. This hypothesis cannot be rejected. Although it can be seen from Fig. 1 that the difference in performance under strong and weak CS was larger under the strong UCS than the weak UCS, this difference was not large enough relative to error variance to be reflected in

TABLE 1

ANALYSES OF VARIANCE OF TOTAL NUMBER OF ANTICIPATORY RESPONSES DURING ACQUISITION FOR THE EIGHT GROUPS ACCORDING TO LATENCY OF RESPONSE

Source	*df*	150–300 Msec. *F*	300–565 Msec. *F*	150–565 Msec. *F*
CS Intensity (CS)	1	13.89*	1.52	7.15*
Replications (R)	1	0.17	0.01	0.01
UCS Intensity (UCS)	1	21.40*	69.42*	77.62*
CS × UCS	1	2.28	0.04	0.71
CS × R	1	0.06	0.01	1.71
UCS × R	1	0.42	0.31	0.02
CS × R × UCS	1	0.10	4.83	1.46
Within groups (error)	152	(62.62)[a]	(202.04)[a]	(309.63)[a]
Total	159			

* $P < .01$.
[a] Error estimate.

a significant CS × UCS interaction. However, a t test of the difference in CS intensity was significant for the strong UCS groups ($t = 2.90$) but was not significant for the weak UCS groups ($t = 1.16$). From an examination of the curves presented in Fig. 1, it appeared that the strong CS–strong UCS group had approached an asymptotic level at about Trial 50. Since the performance of the other groups had not leveled off at Trial 80, this would have the effect of reducing the CS × UCS interaction since the strong and weak CS groups under the strong UCS would be closer together from Trials 50 to 80 than the strong and weak CS intensity groups under the weak UCS. It was therefore decided to perform a *post hoc* analysis of the CS × UCS interaction over the first 50 trials only. The within-groups variance was used as a basis for evaluating the mean difference (Walker & Lev, 1953), and since the direction of the difference had been predicted a one-sided hypothesis was tested. This test resulted in an interaction value which was not significant ($t = 1.28$, $P = .10$), but which was in the predicted direction, with the difference in performance under CS intensity being greater for the strong than the weak UCS groups.

The groups in which CS intensity was shifted during extinction may be regarded as replications of the CS and UCS conditions during acquisition. The small variance attributable to replication indicates that the random assignment to groups was successful.

A separate analysis of variance was performed on responses falling in the interval from 300 to 565 msec. after CS onset. Spence (1953) has called such longer latency responses "CRs" to differentiate them from the short latency "voluntary" responses. The rationale for the separate analysis of conditioned and "voluntary" responses is evidence from a previous experiment (Taylor, 1954) that these responses may be subject to different laws. In that study, although frequency of "voluntary" responses appeared to be a function of CS intensity, frequency of the longer latency "CRs" did not. Summary of the analysis of "CRs" from the present study is presented in Table 1. Bartlett's test of homogeneity of variance yielded a nonsignificant chi square value. It can be seen from Table 1 that UCS intensity was the important variable, CS intensity having no reliable effect on "CRs."

In Table 1, the results of the analysis of the shorter latency, "voluntary" responses are also presented. Although CS intensity as well as UCS intensity produced significant variance, the assumption of homogeneity of variance was violated. However, recent studies (Lindquist, 1953) have indicated that heterogeneity of variance does not substantially alter the outcome of the analysis of variance test.

Because CS intensity reliably influenced the short latency responses, but not the longer latency responses, an analysis of the form of the responses in the two intervals was made to determine whether the short-interval responses were primarily voluntary in form. The voluntary response shows sharp, smooth closure which is maintained until after the onset of the UCS, whereas the form of the CR is irregular. Table 2 presents the mean number of voluntary and conditioned responses differentiated by form in the two intervals. The responses with voluntary form were fairly evenly distributed between the short and longer latency

TABLE 2

MEAN NUMBER OF ANTICIPATORY RESPONSES DURING ACQUISITION FOR THE FOUR CONDITIONS ACCORDING TO FORM AND LATENCY OF RESPONSE

Condition		Voluntary			CR		
CS	UCS	150–300 Msec.	300–565 Msec.	150–565 Msec.	150–300 Msec.	300–565 Msec.	150–565 Msec.
Weak	Weak	1.45	0.75	2.20	2.50	22.78	25.28
Weak	Strong	3.25	3.48	6.73	4.60	38.33	42.93
Strong	Weak	2.20	0.58	2.78	4.57	25.23	29.80
Strong	Strong	7.03	6.70	13.73	7.37	38.33	45.70

intervals, with 55% in the short, and 45% in the longer interval. However, for this population, they constituted a much higher proportion of the total responses in the short latency interval (42.2%), than in the longer interval (8.4%).

Separate analyses of variance were done on the responses differentiated by both form and latency. Table 3 summarized the results of these analyses. The voluntary form responses in both intervals were affected by UCS intensity, however only the short latency voluntary responses were influenced by CS intensity. The interaction between CS and UCS intensity was significant at the .05 level for voluntary form responses in both intervals. UCS intensity influenced the CRs in both intervals, but CS intensity affected only the short latency CRs. There was no tendency toward an interaction between CS and UCS intensity for the CRs in either interval.

Extinction of anticipatory response.— The performance curves showing the mean percentage of anticipatory response per block of 6 trials during extinction appear in Fig. 2. Responses occurring in the interval 150 to 2500 msec. after CS onset were considered anticipatory. It can be seen that the curves for the various CS intensity groups under the weak UCS show little trend. The expected decrease in the percentage of anticipatory responses with increased trials is not clearly evident. Only Groups 1 and 3 show a decrement in performance from the first to last extinction

TABLE 3

ANALYSES OF VARIANCE OF TOTAL NUMBER OF ANTICIPATORY RESPONSES DURING ACQUISITION FOR THE FOUR CONDITIONS ACCORDING TO FORM AND LATENCY OF RESPONSE

Source	*df*	150–300 Msec.		300–565 Msec.		150–565 Msec.	
		Voluntary *F*	CR *F*	Voluntary *F*	CR *F*	Voluntary *F*	CR *F*
CS Intensity (CS)	1	9.91*	8.37*	3.84	0.31	8.74*	2.11
UCS Intensity (UCS)	1	21.24*	8.54*	32.36*	42.90*	36.48*	44.67*
CS × UCS	1	4.43	0.14	4.78	0.31	6.29	0.12
Within groups (error)	156	(20.66)[a]	(28.11)[a]	(24.20)[a]	(191.32)[a]	(65.64)[a]	(251.98)[a]
Total	159						

* $P < .01$.
[a] Error estimate.

trial, with Groups 2 and 4 showing little change in performance over the 30 trials.

In Fig. 2, the performance curves for the strong UCS condition are also presented. Groups 5 and 7 show fairly orderly extinction curves. However, Groups 6 and 8 show little change in performance as a function of increased extinction trials. It can be seen that the strong UCS had the effect of increasing the number of anticipatory responses during extinction, since all groups under the strong UCS are superior in performance to those under the weak UCS.

The results of an analysis of variance performed on the number of anticipatory responses in the interval 150 to 2500 msec. after CS onset during extinction are summarized in Table 4. Analysis of the between-groups variance showed that UCS intensity was the only reliable determinant of response strength during extinction, with CS intensity having no significant effect on performance.

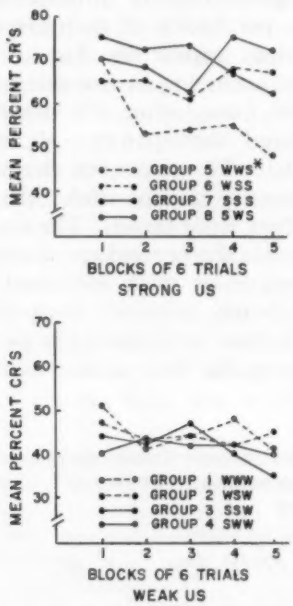

FIG. 2. Mean percentage of responses in the interval 150 to 2500 msec. during extinction per block of 6 trials as a function of CS and UCS intensity. In the legend, the first letter stands for CS intensity during acquisition (W = weak, S = strong). The second letter stands for CS intensity during extinction. The third letter stands for UCS intensity.

TABLE 4

ANALYSIS OF VARIANCE OF TOTAL NUMBER OF ANTICIPATORY RESPONSES IN THE INTERVAL 150 TO 2500 MSEC. AFTER CS ONSET DURING EXTINCTION FOR THE EIGHT GROUPS

Source	*df*	*MS*	*F*
CS Intensity, Acquisition (CS-A)	1	37.06	
CS Intensity, Extinction (CS-E)	1	18.91	
UCS Intensity (UCS)	1	1735.81	16.70*
CS-A × UCS	1	127.81	1.23
CS-A × CS-E	1	12.66	
UCS × CS-E	1	23.26	
CS-A × CS-E × UCS	1	24.78	
Within groups (error)	152	103.93	
Total	159		

* $P < .01$.

A simple analysis of variance was done using the same criterion for an anticipatory response as that used during acquisition, i.e., any response which occurred in the interval 150 to 565 msec. after CS onset. Since the between-groups variance was not significant, no further analysis of these data was undertaken.

DISCUSSION

Acquisition.—The acquisition results of this experiment have provided clear-cut evidence of a positive effect of CS intensity on performance and are therefore consistent with a part of Hull's postulate of stimulus intensity dynamism. In eyelid conditioning, CS intensity has not previously been shown to influence response strength. It was pointed out earlier that the negative findings of these studies may have been related to the level of the UCS employed. In this experiment, the main hypothesis that CS intensity would have a greater effect on performance under strong UCS intensity than under weak UCS

intensity was not clearly verified since the CS × UCS interaction was not significant. However, the fact that the difference between strong and weak CS intensity was significant under the strong UCS but was not reliable under the weak UCS is at least consistent with the hypothesis. Further evidence that UCS intensity may be one of the conditions which influences the effectiveness of CS intensity on behavior comes from the results of the analysis based on the first 50 trials. Although significant at only the .10 level, this result at least suggests that the interaction between CS and UCS intensity may be a real effect.

It may be seen from Table 1 that response strength is more determined by UCS intensity than CS intensity. The finding that a strong level of UCS reliably facilitated performance is consistent with previous studies (Passey, 1948; Spence, 1953; Spence & Taylor, 1951) which have shown UCS intensity to influence acquisition of the conditioned eyelid response. Since the effect of CS intensity on performance is relatively small, a greater number of *S*s is required in order to demonstrate it than in the case of UCS intensity. This may be related to the fact that the results of studies on UCS intensity in eyelid conditioning have been more consistently positive than those on CS intensity.

The analysis of responses by form showed that the tendency toward an interaction between CS and UCS intensity appeared to be due almost entirely to the voluntary form responses, and did not seem to depend on latency. The analysis also showed that for CS intensity, the important factor is the latency of the response rather than the form. CS intensity influenced both voluntary form and CRs in the short latency, but not the longer latency interval. The finding that CS intensity affected the short latency responses but not the longer latency responses is in agreement with the results of a previous study of the eyelid response (Taylor, 1954). Further evidence that the short and long latency responses may be subject to different laws is the finding that UCS intensity had a reliably greater effect on the longer latency responses than the short latency responses. A 2 × 2 × 2 factorial analysis with UCS, CS, and latency as variables yielded significant values for the three main variables and a significant UCS × Latency interaction ($P < .01$). The short latency responses did, however, behave like the longer latency responses in showing a significant increment during acquisition under the strong CS-UCS condition as indicated by an analysis of trend (Lindquist, 1947). It is difficult to evaluate these results since there has been no systematic study of the variables which influence the differentiated form and latency responses. The results of the present experiment, however, suggest that such an investigation may prove worthwhile.

Extinction.—Since the extinction results showed that CS intensity during acquisition had no differential effect on performance, this indicated that there was no reliable effect on learning or habit strength. This interpretation is possible because the factorial design used ruled out differential effects of CS intensity on response strength. The present experiment, therefore, offers no support for the part of Hull's stimulus intensity dynamism postulate (V_1) which specifies that CS intensity affects habit strength.

Although the acquisition results supported the stimulus dynamism prediction that CS intensity would affect performance or reaction potential, the extinction results did not, since the variance associated with CS intensity during extinction was not significant. It is difficult to explain why CS intensity should affect acquisition but not extinction of a CR. Kessen (1953), who reported a similar finding in an instrumental conditioning situation, suggested that the decreased drive resulting from omission of the UCS during extinction could have been a factor. In this experiment, however, such an explanation is inadequate. The UCS was presented on all extinction trials and that the drive level was main-

tained can be inferred from the significant difference in performance under strong and weak UCS during extinction.

The gradual, continuous decrease in performance which McAllister (1953) obtained was shown only by the two groups under the strong UCS in which CS intensity remained the same from acquisition to extinction. Since the two groups in which CS intensity was shifted failed to show the expected decrement it may be that shifting CS intensity reduced extinction. Another possibility is that the gradual response decrement did not appear for the shifted groups because of a depressed response level at the outset of extinction associated with stimulus generalization. On the first extinction trial the shifted groups were reliably inferior in performance to the nonshifted groups.

The lack of a general decreasing trend in performance for the groups under the weak UCS is understandable in view of McAllister's finding that for the 2500-msec. interval employed there is a random response rate of about 40%. Since in the present study the groups under the weak UCS began extinction with low performance levels of 40–50%, there was little opportunity for extinction in these groups. Under the strong UCS condition the groups showed a response level considerably higher at the end of extinction than the 40% random rate obtained in the McAllister study. There was a methodological difference between the two studies since McAllister used a partial reinforcement procedure with 20% test trials during acquisition and extinction where the UCS was omitted, whereas in the present study the UCS was presented on all trials. It seems unlikely that this could have accounted for the higher final response level in the strong UCS groups in the present study since partial reinforcement has consistently increased resistance to extinction (Jenkins & Stanley, 1950). It appears more reasonable that the higher performance was associated with increased drive resulting from the stronger UCS employed in the present study.

Although the acquisition results of this experiment were consistent with the stimulus intensity dynamism postulate, the extinction results offered little support for it. To the extent that the finding of a CS intensity effect on performance during acquisition could not be duplicated during extinction, it would appear that the extinction data may not have provided a fair basis for evaluating that part of the hypothesis which concerns habit strength.

Summary

The present experiment investigated the effect of CS intensity on acquisition and extinction of the conditioned eyelid response under different levels of UCS.

Eight groups of 20 male *Ss* each, were given 80 acquisition trials under four different stimulus conditions: weak CS–weak UCS; strong CS–weak UCS; weak CS–strong UCS; strong CS–strong UCS, with two groups under each condition. Each group was then administered 30 extinction trials with the CS-UCS interval shifted from 500 to 2500 msec. All groups received the same UCS intensity as during acquisition with four groups receiving the same CS intensity as in acquisition and the other four groups being shifted from either weak to strong, or strong to weak CS intensity.

The results showed that CS intensity reliably influenced response strength during acquisition. CS intensity affected the short latency, voluntary form, and conditioned responses, but not the longer latency responses. The latter, however, were more influenced by UCS intensity than the former. There was some evidence that CS intensity had a more pronounced effect on performance under a strong UCS than under a weak UCS. This effect was restricted to the voluntary form responses. The extinction results showed no evidence of an effect of CS intensity on either learning or performance. However, performance during both acquisition and extinction was a function of UCS intensity.

In general, the results of this experiment supported the part of Hull's stimulus dynamism principle which postulates an effect of CS intensity on response strength. There was, however, no evidence of an effect of CS intensity on learning as distinct from performance.

REFERENCES

BARNES, G. W. Conditioned stimulus intensity and temporal factors in spaced-trial classical conditioning. *J. exp. Psychol.*, 1956, **51**, 192–198.

BROWN, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.

CARTER, L. F. Intensity of conditioned stimulus and rate of conditioning. *J. exp. Psychol.*, 1941, **28**, 481–490.

GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: I. The conditioned eyelid response to light. *J. exp. Psychol.*, 1948, **38**, 690–696.

GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: II. The conditioned galvanic skin response to an auditory stimulus. *J. exp. Psychol.*, 1949, **39**, 35–40.

HOVLAND, C. I. The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279–290.

HULL, C. L. Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.

HULL, C. L. *Essentials of behavior*. New Haven: Yale Univer. Press, 1951.

JENKINS, W. O., & STANLEY, J. C., JR. Partial reinforcement: A review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.

KESSEN, W. Response strength and conditioned stimulus intensity. *J. exp. Psychol.*, 1953, **45**, 82–86.

LINDQUIST, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953. Pp. 78–90.

MCALLISTER, W. R. The effect on eyelid conditioning of shifting the CS-US interval *J. exp. Psychol.*, 1953, **45**, 423–428.

PASSEY, G. E. The influence of intensity of unconditioned stimulus upon acquisition of a conditioned response. *J. exp. Psychol.*, 1948, **38**, 420–428.

PAVLOV, I. P. *Conditioned reflexes*. London: Oxford Univer. Press, 1927.

RAZRAN, G. The dominance-contiguity theory of the acquisition of classical conditioning. *Psychol. Bull.*, 1957, **54**, 1–46.

SPENCE, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

SPENCE, K. W., & TAYLOR, J. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

TAYLOR, E. Generalization of the conditioned eyelid response to an auditory stimulus varying in intensity. Unpublished doctoral dissertation, Univer. of Iowa, 1954.

WALKER, H. M., & LEV, J. *Statistical inference*. New York: Holt, 1953. Pp. 158–160.

(Received June 3, 1959)

Journal of Experimental Psychology
Vol. 59, No. 5, 1960

TRACKING PERFORMANCE ON A SEQUENCE OF STEP FUNCTIONS WHICH APPROACHES A CONTINUOUS FUNCTION AS A LIMIT[1]

DAVID McCONNELL AND MAYNARD W. SHELLY[2]

Ohio State University

A number of previous studies (Battig, Gregg, Nagel, Small, & Brogden, 1954; Battig, Voss, & Brogden, 1955; Craig, 1949; Craik, 1948, Ellson, Harris, & Craig, 1949; Ellson & Hill, 1948; Ellson & Wheeler, 1949; Hartman & Fitts, 1955; Searle & Taylor, 1948) have undertaken to relate stimulus amplitude and response amplitude in tracking tasks. In particular, the Battig studies investigated the effect of an increase in frequence of target intermittency on tracking error and found in the later study (Battig, Voss, & Brogden, 1955) a linear relation between this increase and proficiency of tracking.

The present study investigates pursuit tracking performance as the discrete movements of a target on a cathode ray oscilloscope (CRO) increase in number and decrease in amplitude. The discrete target movements approach the limiting continuous movement in a manner analogous to a discrete function $D_n(x)$ uniformly converging to some continuous function $C(x)$, where n represents the number of steps, and x the displacement of the target with respect to the center of the display.

Two important differences between the Battig et al. studies and our own, which affect the extent to which results may be compared, are: (*a*) the use of a step function rather than an intermittent function does not entail possible influences on the response by factors such as brightness or stimulus perseveration associated with intermittent stimuli, and (*b*) the present experiment systematically varies the frequency of the underlying target and also includes a random target condition in order to provide more information about the generality of the relation between stimulus discontinuity and tracking error. In comparing the Battig et al. *limiting* target, which was dependent for perceptual continuity on a rate of intermittency in excess of the fusion rate, with our continuous target, there is little or no reason to believe that these two modes of presentation would, in and of themselves, affect tracking error differently.

METHOD

Apparatus.—The apparatus used in this study is depicted in Fig. 1. The continuous function generator consisted either of a Hewlett-Packard variable-frequency oscillator or of a gas-tube random signal generator with mean frequency of 10 cpm. Under *continuous* conditions the output of the function generator was fed directly to the mixer, and from thence to the display and scoring circuits, while under *discrete* conditions it was introduced into the quantizer, which broke the signal up into particular amplitudes and distributions of steps. Components including

[1] This research was carried out in the Laboratory of Aviation Psychology and was supported in part by the United States Air Force under Contract AF 41(657)-70 with the Ohio State University Research Foundation, monitored by the Aero Space Medical Laboratory, Wright Air Development Center. Permission is granted for reproduction, publication, use, and disposal in whole and in part by or for the United States Government. The aid of Paul M. Fitts, now at the University of Michigan, is gratefully acknowledged by the authors.

[2] Now with the Office of Naval Research, Chicago, Illinois.

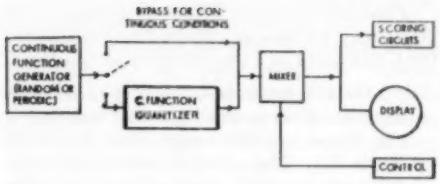

FIG. 1. Flow diagram of the apparatus.

and to the right of the mixer in Fig. 1 are those of Electronic Pursuit Apparatus (EPA) III, an earlier version of which has been described in detail elsewhere (Warren, Fontaine, & Clark, 1952). In the present application EPA was used to present the target and cursor signals on a 5-in. cathode ray oscilloscope (CRO) and to generate various measures of the discrepancy (error) between the two signals. The target on the CRO appeared as a line 18 mm. long and .5 mm. wide, and moved horizontally across the face of the scope. The top of the cursor line was displaced 10 mm. below the top of the target line and was of similar dimensions. Together, when in alignment, they formed a line 28 mm. long. The surface of the CRO was about 28 in. from *S*, 10° below eye level, and perpendicular to the line of sight. The target and cursor were painted on the display by a 200-cps time-sharing switch. Except for large discrete jumps, this produced no stroboscopic effects.

The room in which *S* served was dimly illuminated. Specular reflection from the CRO was eliminated by means of Polaroid filters. The *S* sat in a chair which could be adjusted for height. The elbow of his right arm rested on the pivot point of a horizontal lever and his fingers grasped a knob at the end of the lever. Attached to the control was a potentiometer. The voltage across this potentiometer determined the position of the cursor on the CRO. A maximum movement of the control 40° to the right or left corresponded to a movement of the cursor on the CRO of 1.5 in. to the right or left, respectively. The arm control was of light construction, and was not spring-centered.

Included in the scoring circuits was an amplitude analyzer (Erdman, Fontaine, & McConnell, 1959) designed to record the amplitude distribution of errors generated by *S* when attempting to track the target. This device breaks a continuous error function (a voltage signal) down into a discrete amplitude distribution by measuring the error voltage at .1-sec. intervals and recording a count, proportional to the amplitude category of this voltage, on a cumulating counter corresponding to that particular amplitude.

Normally, this apparatus recorded errors to the left and to the right of the target in appropriate counters. However, it was possible to throw a switch so that the apparatus would record errors which were dependent upon the direction of the target for all signals which did not change direction between points of maximum excursion. When

TABLE 1

DESIGN OF THE EXPERIMENT

Number of Steps into Which Range is Divided	Distribution of Discrete Step Inputs	Frequency of Signal			
		Triangular			Random
		10 cpm	20 cpm	30 cpm	10 cpm
2	Equal	T-2-0-10	T-2-0-20	T-2-0-30	R-2-0-10
4	1	T-4-1-10	T-4-1-20	T-4-1-30	R-4-1-10
	2	T-4-2-10	T-4-2-20	T-4-2-30	R-4-2-10
	3	T-4-3-10	T-4-3-20	T-4-3-30	R-4-3-10
8	1	T-8-1-10	T-8-1-20	T-8-1-30	R-8-1-10
	2	T-8-2-10	T-8-2-20	T-8-2-30	R-8-2-10
	3	T-8-3-10	T-8-3-20	T-8-3-30	R-8-3-10
12	Equal	T-12-0-10	T-12-0-20	T-12-0-30	R-12-0-10
Continuous	—	T-C-0-10	T-C-0-20	T-C-0-30	R-C-0-10

this switch was in the lead–lag position, it recorded counts in one band of counters when *S* was ahead of (led) the target, and in another bank of counters when *S* was behind (lagged) the target. On odd-numbered trials this switch was in the lead–lag position, and on even trials in the left–right position.

Subjects.—Five naive *S*s were used. All were male, right-handed students who received hourly wages.

Design.—The experiment consisted of 36 conditions. All *S*s served six times in each condition. The 36 conditions are presented in Table 1.

As indicated, two types of underlying signals were used—a random signal and a triangular signal. The triangular signal was used because the average increment of its corresponding step input is linearly related to the theoretical error between continuous and discrete signals. The random signal had a normal distribution of amplitudes and its power spectral density was peaked at 10 cpm. The frequencies of the triangular signal were 10, 20, and 30 cpm. The range of both functions was divided into 2, 4, 8, and 12 discrete steps, or was continuous. For 4- and 8-step functions, one of three distributions of the steps (relative to the mean increment size) was used. The sets of numbers and letters in the cells of the design (Table 1) give the code by which the individual conditions were designated. In each cell, the first position refers to the wave form ("T" for triangular, and "R" for random). The second position refers to the number of steps into which the signal is quantized. When a continuous function is used, a "C" stands in the second place. The third position refers to the particular distribution of increments used in quantizing the function. If all the increments were equal, an "O" is in this position. The last position refers to the frequency of the underlying function in cycles per minute. Thus, for example, T-8-2-20 refers to the triangular eight-step, second distribution, 20-cpm function. In the bottom row of Table 1, the distribution index (0) can be ignored since the functions are continuous.

In the distribution of the amplitudes of the step inputs for all numbers of discrete inputs used in this study, the same maximum excursion was maintained. The amplitudes of the step inputs were all equal for the 2-step and 12-step conditions. For the 4- and 8-step conditions, however, a variety of distributions was used.

Procedure.—When *S* first reported to the laboratory he was seated on *S*'s stand facing the CRO, and given the following instructions:

> On the scope in front of you are two signals. One is moving. We will call it the target. (The target was T-C-0-10, Table 1.) The other is stationary. We will call it the cursor. Now take the handle in your right hand. Make your arm comfortable. Now note that you can move the cursor by moving the arm control. (*E* moved *S*'s arm back and forth to demonstrate, making a rough effort to track the target. After a few seconds, he removed *S*'s hand from the control.) You will see several different kinds of targets: some will be chopped up and some will move smoothly; some will move faster than others. Your job is to keep the horizontal distance between the target and cursor at a minimum at all times. Each period will be just over a minute long, and at the start of each period. I will say 'ready,' from the next room, and if you say 'yes,' we will begin. Any questions?

Each trial was of 65-sec. duration, the first 5 sec. being an unscored warm-up period. Between trials *S* rested 2 min., except that he was permitted two 10-min. breaks in a day's run. Each *S* served for an average of 27 trials each day, or a little less than 2 hr., including rest periods. This continued for eight days. During the experiment each *S* went through each condition six times. The order each time was random with the obvious restriction that conditions were not repeated until *S* had served in all 36 conditions an equal number of times.

Measures.—The theoretical error between the periodic step function and the continuous function was derived as follows: The distance between the step function and the continuous function was squared and its definite integral with respect to time computed. After averaging, the square root was then taken. The result is symbolized by $E = (\mu_2)^{\frac{1}{2}}$ where μ_2 is the usual symbol for the second moment about the origin which, in this case, was taken as the continuous signal.

On every trial two voltage readings were taken. The first represented the voltage difference between the cursor and the target (T) squared and integrated over the last 60 sec. of each trial. The second reading represented the squared and integrated voltage difference between the cursor and the underlying continuous function (C). When *S* was tracking the continuous function the two

voltage readings should have been the same, but because of limitations of the scoring circuits it was not possible to maintain this condition for all input voltages. Consequently, a curve relating the two circuits was plotted as a function of input voltage, and from it the two voltmeter readings were adjusted so that the scale factors were the same for both scores. This curve was linear over most of the range.

The first score is represented by E_T and the second one by E_C. Symbolically, if the difference between the signal and cursor at Time t is e_T or e_C, the condition (T or C) and time interval over which the score is integrated by I_T or I_C, then:

$$E_T = \left[\frac{\int_I e_T^2(t)dt}{I_T}\right]^{\frac{1}{2}} = (\mu_2,C)^{\frac{1}{2}} \quad [1]$$

where t stands for time. Similarly,

$$E_C = \left[\frac{\int_I e_C^2(t)dt}{I_C}\right]^{\frac{1}{2}} = (\mu_2,C)^{\frac{1}{2}} \quad [2]$$

Note that neither E_T nor E_C is strictly rms error *unless* *S*'s error is centered about the target.

The two voltage readings were then converted to inches measured on the face of the scope. This was possible through a knowledge of the constants of integration for the meters. Such a conversion of scores was linear. As a further check on the accuracy of the conversion, the converted scores were compared with similarly defined scores computed on the basis of the amplitude analyzer data. The average absolute difference between the converted meter scores and the scores based on the amplitude analyzer for a sample of scores was 4.57%. The average error of comparison for this same sample was .39%.

Amplitude analyzer data were collected on every trial. For the random function this entailed the recording of errors to the left or right. For the triangular function, on even-numbered presentations of the conditions, left–right errors were recorded, and on odd-numbered presentations of the conditions, lead–lag data were collected. Graphic records were taken periodically of target, cursor, and error voltages in an effort to determine at what number of steps (if any) *S* changed from one mode of responding (i.e., continuous or discrete) to the other.

Results

Although measures were taken throughout the six cycles, results reported are based only on the last five complete cycles when *S* was highly practiced. Because of this, the error term was small enough to permit the use of parametric procedures when needed. Since, however, Wilcoxon's paired replicates test was appropriate to many situations in which it was desired to test differences, this nonparametric test was the most extensively used. As too many of the scores for R-2-0-10 (the 2-step condition with a random underlying signal) reached the maximum of the scoring devices, the scores obtained for that condition seriously underestimate the true scores. This was not true of any other conditions. Therefore, R-2-0-10 was not included in the analyses. Whenever a result is referred to as significant, the .05 level may be assumed.

Figure 2 shows the average error (E_T) with respect to target as a function of the target step size (θ). Results are very similar for the error (E_C) with respect to the underlying continuous signal. In the figure the three distributions of step input amplitudes have been averaged for both 4- and 8-step conditions. Because of the concern in this study with the approach to a limit, it is relevant to raise the question as to whether or not the findings are specific to the type of distributions used in these conditions. The conclusion seems to be negative on several grounds. First, it might be suspected that an even distribution, because of its closer approximation (smaller E) to the continuous triangular function, would lead to smaller error scores. Approximation to the underlying continuous function did not, however, linearly order the resulting error on the different distributions. This held for both E_T and E_C. Furthermore, using Wilcoxon's paired repli-

cates test and considering only those pairs which have the same underlying continuous distribution, of the 24 pairs which were tested only two showed significant differences. These were both for 4-step conditions. Furthermore, with the data upon which Fig. 2 is based, when linear functions were fitted to the data points without first averaging the values obtained for different distributions, no significant differences were obtained between the obtained slope or intercept estimates when Wilcoxon's paired replicates test was used. It will also be recalled that the data underwent a number of transformations before being fitted with linear functions. Only one of these for one meter, however, could have altered the values so that they could be better described by a linear function. The unaltered scores obtained from this meter (for E_C) were plotted and fitted with linear functions. As was found with the transformed scores, the data points could be described by linear functions with the error substantially between 2% and 3%. Thus there are a number of strong reasons for believing that the functions obtained are estimates of general relation, and are *not* dependent upon the precise distribution of steps in the discrete function.

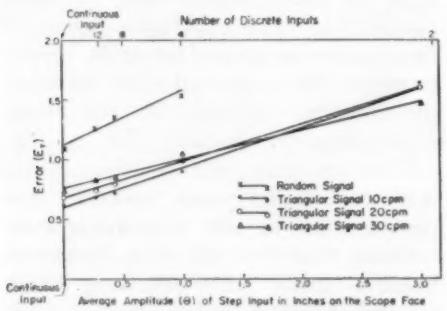

FIG. 2. Average error (E_T) with respect to target as a function of step input amplitude (θ).

In Fig. 2 all of the sets of points corresponding to a single underlying continuous signal have been fitted with linear functions. A least-squares technique was not used for two reasons. It was felt that some of the extreme deviations could as easily be the result of apparatus malfunctioning as a result of the behavior of *Ss*. Data immediately preceding any apparatus failure were not included, but this precaution was undoubtedly not foolproof. Second, it was felt that since curves were being fitted to the data of individual *Ss* before being averaged, the least-squares method would add unduly to the variance and affect the final estimates somewhat more. Therefore, the method of successive differences was used.

Two points emerge from inspection of Fig. 2. The first is that both the slope and intercept of the function describing performance on the random signal conditions are greater than these constants for the triangular conditions. This difference was substantiated by Wilcoxon's paired replicates test, using trials as the basis for matching scores. In both cases, that triangular function which led to performance closest to that on the random conditions was used for comparison.

The second point is that the slope of the function fitting the 10-cpm triangular signal data is greater than the slope of the function fitting the 30-cpm data. This difference was tested in a similar way and found to be significant. Since these two functions intersect at 1.75-in. step amplitude, it follows that opposite relations between frequency of the underlying continuous signal and error obtain to the left and to the right of the intersection; that is, for lower amplitudes, increasing signal frequency from 10-cpm to 30-cpm increased

error, while for higher amplitudes the frequency increase resulted in lower error.

The distribution of errors.—The distributions of errors obtained from the amplitude analyzer indicated two things: (*a*) that for all conditions *S*s were on the average to the left of the target, and (*b*) that for all triangular conditions *S*s were leading the target with respect to its prevailing direction (Fig. 3). Lead and lag information was not determined for the random functions because of an instrument limitation. Significance levels indicated on Fig. 3 were determined by conducting sign tests on the equality of distribution frequencies in certain critical zones of counters. A dash adjacent to a function indicates that the effect of increasing the number of steps was nonsignificant at the .05 level. The two significance levels indicated by vertical arrows are associated with similar sign tests to detect distributional differences attributable to frequency for these functions. The conversion from the distributional frequencies to inches on the CRO was accomplished by reducing the apparent mode of a particular distribution to a counter number, then referring to a conversion table computed from previous calibration data. When distributions were multimodal, they were smoothed by fitting a polynomial, and the apparent mode was then taken as the counter number nearest to the curve maximum.

Figure 3 shows that the mean amount of lead increases with the number of steps, but the levels of significance between adjacent increments make it difficult to ascribe a single uniform function to this relationship. In fact, it would appear that for relatively low step numbers, frequency does not order the amount

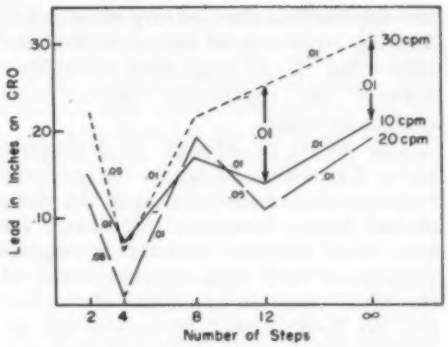

FIG. 3. Lead on the CRO display as a function of the number of steps. The decimals adjacent to the functions indicate the outcome of sign tests of the distributional differences in selected groups of counters.

of lead, while for relatively high step numbers (including the continuous case) there is a significant effect of frequency on lead. On the other hand, the number of steps appears to exert a significant effect on the amount of lead, almost regardless of frequency. In particular, the 4-step function appears to be associated with the smallest lead.

A remarkably consistent relationship appeared in comparison of error distributions for 4-, 8-, and 12-step random and triangular step functions. For the 2-step and continuous functions, consistent differences between the two types of functions did not appear. For the others the random function clearly shifted errors to the left of the target—significantly further to the left than for the triangular function.

DISCUSSION

The findings indicate that over the sequence of experimental conditions the relationship between average amplitude of the step input and the particular error measure selected was linear. Put another way, as the sequence of step inputs more closely approximated the underlying continuous signal, the error meas-

ure approached the limiting error. The limiting error was, of course, well above zero—that is, *S*s responded to neither discrete nor continuous targets with perfect accuracy.

The effect of change in frequency upon *S*'s ability to track the periodic function was merely to change the slope of the linear function. Moreover, for two very diverse underlying signals (triangular and random) the form of the error function remained the same; i.e., *the form of the approach to the behavioral limit within these conditions was independent of the character of the limiting signal.* Thus, for the measure used, the extent to which *S* could track a discontinuous signal could be predicted simply by a knowledge of his performance on the corresponding continuous signal and the slope constant of the linear function. Moreover, this common property to both periodic and random signals emerges despite the unsurprising fact that use of the random input clearly degrades performance with respect to the periodic input.

There is evidence, e.g., from graphic records, that even at eight steps with the high-frequency periodic signal *S*s make distinct responses. Yet not until the number of steps is reduced to four, does the particular distribution of a fixed number of steps across the tracking range appear to have any differential influence on error. This means that *S* can compensate without degradation of his response, for variation in the positions of discrete stimuli along the underlying path, up to some limit (here, apparently, four steps). What the determinants are of this "limit," beyond which *S* loses ability to "smooth" the function equally well for all distributions, is a question open to speculation. Taylor (1948) has stated that he believes *S*s develop sets for responding to the average size of error (average amplitude) and that this influences all of his responses in the direction of the average. A number of later investigations (Craig, 1949; Ellson & Hill, 1948; Ellson & Wheeler, 1949) have confirmed this, one of them (Ellson & Wheeler, 1949) showing that the effect is due to the range and not to the absolute amplitudes of the step inputs. These findings would seem to indicate that *S*s, by overshooting small amplitude inputs and undershooting large amplitude inputs, would increase their error. The Taylor theory appears to be confirmed in the case of the 4-step function. The larger errors were found, in our experiment, in both cases for the more irregular distribution—where the signal appeared farther from the average signal, and also from the corresponding position of the underlying signal.

In an effort to correlate the lead–lag errors of Fig. 3 with the integrated error of Fig. 2, we offer the postulate that for small to moderate ranges of increment size, *S* minimizes his errors as target difficulty increases by increasing his lead on the target. Under this general hypothesis, the following points arise:

1. The fact that at two steps error is greater than at four, while lead is also greater, can be attributed to the fact that between the two extremes of the 2-step function there were almost 100° of arm movement, which meant that to keep zero error, *S* would have had to stay on the target at one extreme, then move instantaneously to the other extreme. Since this is impossible *S* is confronted with two alternatives—either to stay with the target until it moves, thereby accepting a lag error with every movement, or to try to anticipate by a fraction of a second when the target will move. That *S* chooses the latter course *in the absence of instructions* as to lead–lag behavior is clear from the data. While his error is maximal under this extreme condition it may very well be that within the condition, the anticipatory or *lead* response is still the most efficient—i.e., produces less error than any other mode of response. An obvious experimental check on this inference is to instruct two groups of *S*s differently: the first to anticipate when the target will move by taking a lead on it just before it moves, as *S*s appear to have done in this study, and the second to stay on the target at all costs until it moves.

2. The decrease in lead from eight steps to twelve for 10- and 20-cycle targets, corresponding to the decrease in error, is together with graphic records interpreted by

Es as demonstrating that the mode of response changes from discrete to continuous across the step change 8 to 12. Because of the deliberate omission of instructions relevant to continuous versus discontinuous responding, it is concluded that the optimal mode at eight steps for the two lower frequencies was still discrete, while it was continuous for the 30-cycle target. At 12 steps, the optimal mode appears to have been continuous with a lessening of lead. At this number of steps, no discrete element could be detected in the graphic record for any *S*, for any frequency.

3. The discrepancy between the 30-cpm function and the 10- and 20-cpm functions at 12-step and continuous values of the argument (Fig. 3) corresponds to the finding already discussed (Fig. 2) that 30-cpm error appears to be slightly greater for these arguments. Since the targets are more difficult to track, *S* evidently increases his lead, and, as above, thereby reduces his error to a minimum for that particular target.

It is obvious, therefore, that explicit confirmation of the hypothesis that *S* improves his performance by leading the target must be contingent upon use of targets of equal difficulty, while constraints are experimentally imposed upon lead and lag behavior.

One further key to the interpretation of the lead results is the singular relationship indicated above that as the target increases in difficulty, *S* apparently stays increasingly to its left. Our hypothesis here is that *S* puts most of his lead into the cursor during flexion. (Right-handed *S*s flex to the left.) This notion is clearly complementary to the theory that leading the target improves performance, since we know that flexion responses are, in general, anatomically easier to control. Since we were unable to make comparisons of lead–lag and left–right distributions on the same trial, the experimental accreditation of this hypothesis is not entirely established.

Summary

The relation between error in tracking a continuous function and error in tracking a discrete function which approaches the continuous function in the limit was examined. Random and periodic targets were presented to five *S*s, and response measures included root-mean-square error between target and cursor, as well as between cursor and the limiting continuous function. Also measured were average lead–lag error of the cursor with respect to the target, and average left–right error of the cursor with respect to the target. Results indicated that as the number of steps is increased, the error measure chosen on the discrete targets linearly approached that associated with the limiting continuous target. This linearity was virtually independent of frequency or whether the limiting function was periodic or random. The *S*s on the average led the target under all conditions, and were also to the left of the target on the average. The relation of these results to the previous finding was interpreted as indicating that *S* minimizes his error by leading the target, and that this is especially pronounced during flexion of the control arm.

References

Battig, W. F., Gregg, L. W., Nagel, E. H., Small, A. M., Jr., & Brogden, W. J. Tracking and frequency of target intermittence. *J. exp. Psychol.*, 1954, **47**, 309–314.

Battig, W. F., Voss, J. F., & Brogden, W. J. Effect of frequency of target intermittence upon tracking. *J. exp. Psychol.*, 1955, **49**, 244–248.

Craig, D. R. Effect of amplitude range on duration of responses to step function displacements. *USAF Tech. Rep.*, 1949, No. 5913.

Craik, K. J. W. Theory of the human operator in control systems: II. Man as an element in a control system. *Brit. J. Psychol.*, 1948, **38**, 142–148.

Ellson, D. G., Harris, H., & Craig, D. R. The interaction of responses to step function stimuli: II. Equal opposed steps of varying amplitude. *USAF Memo. Rep.*, 1948, No. MCREXD-694-2P.

Ellson, D. G., & Hill, H. The interaction of responses to step function stimuli: I. Opposed steps of constant amplitude. *USAF Memo. Rep.*, Nov. 1948, No. MCREXD 604-2P.

Ellson, D. G., & Wheeler, L., Jr. The range effect. *USAF Tech. Rep.*, 1949, No. 5813.

Erdman, A. C., Fontaine, A. B., & McConnell, D. G. The amplitude analyzer. *USAF WADC Tech. Rep.*, July 1959, No. 59–45.

HARTMAN, B. O., & FITTS, P. M. Relation of stimulus and response amplitude to tracking performance. *J. exp. Psychol.*, 1955, **49**, 82–92.

SEARLE, L. V., & TAYLOR, F. V. Studies of tracking behavior: I. Rate and time characteristics of single corrective movements. *J. exp. Psychol.*, 1948, **38**, 615–631.

TAYLOR, F. V. Certain characteristics of the human servo. AIEE Misc. Paper 48–235, 1948.

WARREN, C. E., FONTAINE, A. B., & CLARK, J. R. A two-dimensional electronic pursuit apparatus. *USAF Hum. Resour. Res. Cent., Res. Bull.*, 1952, No. 52–26.

(Received June 3, 1959)

Journal of Experimental Psychology
Vol. 59, No. 5, 1960

MULTIDIMENSIONAL MAPPING OF MUNSELL COLORS VARYING IN HUE AND CHROMA[1]

TAROW INDOW AND TSUKIKO UCHIZONO

Keio University, Tokyo

The purpose of this study was to make an experimental inquiry, by the use of the multidimensional scaling, into the possibility that differences perceived among colors would be adequately represented by distances in a real Euclidean space. Here, by color differences, not only ones slightly above the difference limen but also supraliminal ones of considerable degree are to be understood.

As is well known, in the Munsell system, a surface color is specified by a point in a solid in which cylindrical coordinates represent, by definition, the principal attributes of colors. Any horizontal section through the solid is defined as a plane of constant lightness (Munsell value), while any vertical plane originating at the achromatic axis represents a plane of constant hue, and a cylindrical section concentric with the axis constitutes a surface of constant saturation (Munsell chroma). The scaling units for the three attributes are defined so that equal distances on each of the hue, chroma, and value scales would represent equal observed visual intervals of that attribute. Then, one question will arise. Suppose that two colors are given which are respectively 5R 6/2 and 5R 4/8 in the Munsell notation. It is obvious that these colors differ by 2 units in value and by 6 units in chroma. It is not clear, however, how to represent the total impression of difference between these colors. Is it given by the magnitude $\sqrt{(w_1 2)^2 + (w_2 6)^2}$? In the reports of the OSA subcommittee (Newhall, 1940; Newhall, Nickerson, & Judd, 1943) which examined the spacing of Munsell colors in detail, it was stated that no attempt was made to evolve a single scaling unit for the entire solid, or to equate the scaling unit for hue, value, and chroma. Hence, the relative size of these units, e.g., w_1 and w_2, is not known, although several tentative definitions of the relative size have been made (Godlove, 1951). Furthermore, it is not self-evident that the distance corresponding to the difference over more than one attribute is Euclidean in nature and also that value and chroma, for example, are to be represented by axes orthogonal to each other.

The Euclidean character of the Munsell solid is tacitly assumed in various formulae to evaluate color differences quantatively (Judd, 1952). These formulae are regarded, in general, as applicable only to differences among more or less similar colors. That means only the differences slightly above the limen have been taken into consideration when experiments were undertaken to test the adequacy of the Euclidean model in the Munsell system. The same is true in experiments for developing a uniform tridimensional color scale. It has been of primary concern in these experiments to examine whether colors are arranged in a tridimensional array so that each color differs from each of its nearest neighbors by

[1] This research was supported by a grant from the Ministry of Education, Japan. The authors are indebted to M. Matthewson for her helpful advice concerning English.

perceptually equal amounts. In short, uniformity of the scale *in the small* only has been the subject of the experiments, and little attention has been paid to uniformity of the scale *in the large.*

Torgerson (1952, 1958) developed the method of multidimensional scaling and applied it to a set of nine Munsell colors all of the same red hue (5R) but differing from each other in value and chroma. For obtaining comparative distances among these colors the complete method of triads was used in his experiment. The experiment was repeated by Messick (1954, 1956) with different procedures for obtaining distances among colors again varying in value and chroma but kept roughly at 5R in hue. The results of the multidimensional mapping were much the same in both experiments and the configuration of colors agreed closely with the Munsell system. Richardson (1938) obtained essentially the same results, though the procedures of multidimensional mapping were simpler than those of Torgerson and Messick. It is to be noted that color differences *in the large* play an important role in the multidimensional scaling. Torgerson and also Messick pooled the responses of a number of *S*s and obtained a common set of distances among colors to which the multidimensional scaling was applied. The close agreement of the final results with the Munsell system would, therefore, hold only in the responses of *S*s as a whole but not necessarily in the responses of each *S*. It is open to question especially because both methods of scaling the distances presuppose heterogeneity within *S*s. That is to say, the existence of individual differences as to what *S*s perceive between a pair of colors is requisite for the procedures. Such individual differences, however, might be qualitative as well as quantitative in nature. Employing the same nine colors as used by Torgerson, Indow and Shiose (1956) constructed the individual multidimensional configuration of the colors for each of three *S*s. The configurations obtained were in good agreement with each other and also with the Munsell system. In this experiment, the nine colors were of the same red (5R) but varied in value and chroma, from 3 to 7 and 4 to 12 respectively. Variation in hue, from 6R to 6YR, was added to that in value and chroma in the experiment carried out by Shiose (1958). Tridimensional configuration of the colors was successfully constructed in agreement with the Munsell system though the range of hue was rather restricted. In the present experiment, 21 colors were used which varied in hue and chroma over the widest possible range. The lightness was kept constant at 5 in Munsell value. As will be discussed later, among the basic postulates of the Munsell system the most questionable one may be that hue is represented by a radius vector and chroma by a concentric circle in a Euclidean plane.

Method

Apparatus and stimuli.—A standard Color i and 21 colors[2] ($j = 1, 2, \ldots, 21$) were presented simultaneously on a rectangular board, 45 × 95 cm. Each of the 21 colors, .7 × 1.0 cm., was pasted on a piece of thick cardboard, 1.3 × 3.1 cm. The remaining part of the card was covered with gray paper (N 5). When the card was handled, the gray part was always picked up with a pincette which was covered by gray cloth. The standard color was also pasted on a thick cardboard, 1.0 cm. wide and 45 cm. long. This was always placed at the left side of the board. The *S* was seated in front of a table on which was the board. Gray cloth was

[2] The colored papers were obtained from Murakami Color Research Laboratory.

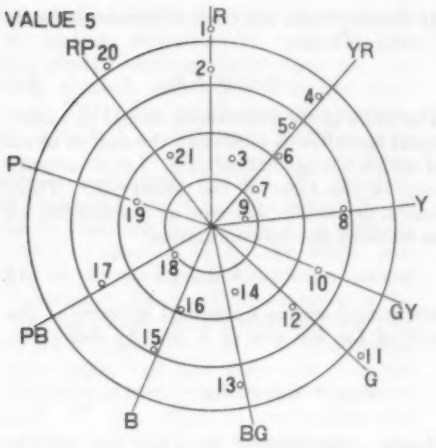

FIG. 1. Color stimuli in terms of Munsell hue and chroma. Concentric circles denote Chroma 2, 4, 6, and 8 respectively. Radius vectors are only for the sake of facilitating identification of colors.

hung from the ceiling around the table. The same gray cloth was pasted on the board to cover its entire surface. Three fluorescent 20-w. daylight tubes (Toshiba FL-SDL) evenly illuminated the surface of the board to 500 lux. Since the spectral energy distribution of this light source was known and the spectral reflectances of the colored papers were measured, it was possible to compute tristimulus values for the colors under this illumination. The Munsell designations were obtained from the tristimulus values through "the table of ICI equivalents of the recommended Munsell notation" (Newhall et al., 1943). Colorimetric data of the gray cloth were as follows; $x = .305$, $y = .302$, $Y = 18.0$, and 3.0P 4.8/1.0. A plot of the stimuli according to hue and chroma appears in Fig. 1 where radius vectors are introduced, just for the sake of convenience, to compare this configuration with those of the final results. In the Munsell system, for instance, 5Y is defined to represent the yellow as a major hue. Nevertheless, the Vector Y in Fig. 1 is not located at 5Y but at approximately 7Y. The Munsell values of the stimuli were between 4.6 and 5.0.

Procedure.—On the gray board, a standard Color *i* (45 cm. long, 1.0 cm. wide) was placed along the edge of the left side (45 cm.). Parallel to the 95-cm.-long side, 21 rows were marked by gray threads on the board, and a Color *j* was placed in each row at its left end. That means, 21 colors made a column next to the standard Color *i* on its right side. The same one as the standard was included among the 21 colors. The *S* was told that this duplicate of the standard would always be present and that he should place it right beside the standard. Then *S* was required to move, one by one, each of the other 20 colors to the right in the row according to the difference he observed between Color *i* and Color *j*. For the purpose of furnishing *S* with a unit for moving the colors, a pair of gray cards, either N 5.0 and N 5.5 or N 5.0 and N 4.5 were placed on the top of the board. The interval between the cards (.7 × 1.0 cm. each) was fixed at 3 cm. Should the difference between this pair of grays be represented by this distance (center to center), how far should the Color *j* be moved to the right to represent the difference between the Colors *i* and *j*? This was what *S* was instructed to do (with full knowledge that distances would be measured from center to center in the strips), hence the method may be called the multiple-ratio judgment. Usually, after having moved all the colors, *S* made subsequent changes in the positions of the colors by surveying the configuration of the colors made on the board as a whole. The final position of each color was thus found by a series of successive approximations. The same color as the standard was always placed by *S* at the left end of the row and from this point the placement were measured in centimeters. Occasionally *S* required to move a color beyond the right end of the board. A supplementary board was added in such cases. Each of the 21 colors took the place of the standard *i* in turn. The spatial order of the 21 colors in the column was randomized from trial to trial.

Subjects.—Four *S*s participated in the experiment. All were normal in color vision. Id is the senior *E* and had served also as *S* in the previous experiment (Indow & Shiose, 1956). On the other hand, K, O, and M had neither experience in an experiment of this kind nor special knowledge about the Munsell system. K and O were women students in the introductory psychology course and M was a housewife. They found, however, no difficulty in converting the color differences into the spatial distances as required, after a few trials. And they were so naive that no special emphasis was necessary to have them take an unanalytical attitude in observing the color differences. As each of the 21 colors took the place of the standard *i*, one round consisted of 21 trials. Id repeated four rounds of the trials and the other *S*s two

rounds, respectively. In a sitting 10 to 14 trials were run.

Scaling procedure of the differences.—The first step of multidimensional scaling is to obtain scale values for the distances between all pairs of stimuli. It is important, however, that the subsequent multidimensional mapping is not independent of the nature of the scale on which the distances are determined. The distances have to be determined, in the sense of Stevens (1951), on a ratio scale and not on an interval scale, because the choice of an arbitrary origin for the scale will affect the dimensionality as well as the nature of the space to account for these interstimulus distances (Messick & Abelson, 1956; Torgerson, 1952, 1958). For obtaining the distances, Torgerson (1952, 1958) employed the complete method of triads, i.e., an extension of the method of paired comparison, and Messick (1954) used the method of successive intervals of a modified form. Either of the methods yields only an interval scale. Hence, they had to convert the relative distances thus obtained to the absolute ones by estimating the value of the additive constant. To choose the correct additive constant is not an easy task and its analytical solution is extremely tedious in computation (Messick & Abelson, 1956). In the previous experiment Indow and Shiose (1956) had to arrive at the absolute distances with the help of the estimated additive constant. Now this difficulty can be overcome by using the method of the present experiment, since the color differences are obtainable from the beginning on a ratio scale.

As the result of the placement of a Color j by S, a spatial Distance x_{ij} is obtained which should represent the observed difference between the Color j and the standard Color i in terms of an arbitrary Unit e_i. The distances x_{ij} ($j = 1, 2, \ldots, n$) may be thought of as being expressed by the same Unit e_i because they are determined at a trial as the results of simultaneous comparisons with each other. However, in general, the Unit e_i will not remain constant if the standard Color i varies. Although a pair of grays are given as a guide for converting the color differences into the spatial distances, as a matter of fact, the proportional constant of the conversion will be more or less affected by the total relations between the standard Color i and the Color j ($j = 1, 2, \ldots, n$). The differences, in general, may be relatively small with a standard color and relatively large with another. In other words, the proportionality will be kept constant among x_{ij}'s in so far as i is fixed, but not necessarily if i varies. Let us denote by δ_{ij} the color difference between i and j. Hence,

$$x_{ij(i)} = \delta_{ij} e_i \qquad [1]$$

The suffix in the parenthesis should be understood hereafter as indicating the unit in terms of which δ is quantified as x. Let us arrange $x_{ij(i)}$ in the form of the Matrix X. Then, taking arbitrarily kth row as a reference, let us consider the transformation

$$x_{ij(k)} = x_{ij(i)} c_{ik} \qquad [2]$$

where $x_{ij(k)}$ means δ_{ij} defined in terms of the unit of the kth row of X and by definition,

$$x_{ij(k)} = \delta_{ij} e_k = \delta_{ij} e_i \frac{e_k}{e_i} = x_{ij(i)} \frac{e_k}{e_i}$$

Hence, remembering $\delta_{ik} = \delta_{ki}$, c_{ik} will be given as follows,

$$c_{ik} = \frac{e_k}{e_i} = \frac{\delta_{ki} e_k}{\delta_{ik} e_i} = \frac{x_{ki(k)}}{x_{ik(i)}} \qquad [3]$$

where $x_{ki(k)}$ denotes the ith element in the kth row and $x_{ik(i)}$ the kth element in the ith row of X. The transformation [2] will be possible in as many ways as n, the number of rows of X. Taking the average

$$\bar{x}_{ij} = \frac{1}{n} \sum_k x_{ij(k)} = x_{ij(i)} \left(\frac{1}{n} \sum_k c_{ik} \right) = x_{ij(i)} \bar{c}_i \qquad [4]$$

($c_{ii} = 1$)

and on account of Equation 3,

$$\bar{c}_i = \frac{1}{n} \sum_k \left(\frac{e_k}{e_i} \right) = \frac{\bar{e}}{e_i} \qquad [5]$$

where $\bar{e}$ is the average of e_i's. From Equations 1, 4, and 5, we have

$$\bar{x}_{ij} = \delta_{ij} \bar{e} \qquad [6]$$

and the whole observed color differences can be defined by a single common Unit $\bar{e}$. Due to random errors, $\bar{x}_{ij}$ will be somewhat different from $\bar{x}_{ji}$. Hence, the mean of $\bar{x}_{ij}$ and $\bar{x}_{ji}$, i.e., d_{ij}, was used through the present study as the value of δ_{ij} on the ratio scale.

As stated previously, Id made four repetitions on each of $x_{ij(i)}$, and the remaining Ss made two repetitions. The average was taken as the element of the Matrix X, and d_{ij} was obtained by the procedures described above for each S respectively.

To see how small the discrepancy is in general between $\bar{x}_{ij}$ and $\bar{x}_{ji}$ will serve as a criterion of consistency in the placements made by S. The correlation coefficients between the two were .98 for Id, .89 for K,

.93 for O, and .85 for M. Besides, the plots of $\hat{x}_{ij}$ with $\hat{x}_{ji}$ showed in all cases that the points plotted were scattered along the line passing through the origin with the slope of one. There was a high degree of consistency in each S to express δ_{ij} as either $x_{i(i)}$ or $x_{j(j)}$. The criterion tells us nothing, however, concerning whether the relationship between x and δ is really as simple as given in Equation 1. Even if δ_{ij} is replaced by $f(\delta_{ij})$, a function of δ_{ij}, the whole procedure of the scaling still holds and, instead of Equation 6, $f(\delta_{ij})\bar{c}$ will be obtained as $\hat{x}_{ij}$. In this case also the criterion will be met since $f(\delta_{ij})$ is equal to $f(\delta_{ji})$. It is a matter of assumption that the observable x is put in Equation 1 as proportional to the latent variable δ. No empirical verification was possible in making this assumption.

The multidimensional scaling was applied to the data of Id and also to the pooled data of K, O, and M. The d_{ij}'s of the three Ss were pooled as follows. In order to reconcile the difference in the Unit $\bar{c}$, the d_{ij}'s of each S were plotted against the d_{ij}'s of Id and the former were multiplied by a constant so that the points plotted were scattered along the line passing through the origin with the slope of one. By taking the unit of the scale of Id as reference, the scales of all Ss were thus defined by a common standard. The converted d_{ij}'s of the three Ss were averaged to give the Common Distance Matrix D.

Multidimensional scaling.—Let i and k be alternative subscripts for stimuli ($i, j = 1, 2, \ldots, n$) and let b_{ij} be the scalar product of the vectors from an origin to Points i and j in a real Euclidean space of r dimensions. As suggested by Torgerson (1952, 1958), it is convenient to place the origin at the centroid of all of the stimuli. Let the scalar product thus defined be denoted by b^*_{ij}, then b^*_{ij} can be derived from Matrix D by the formulae given by Torgerson. Once Matrix B^*, whose elements are b^*_{ij}, is obtained, it becomes possible to determine the dimensionality of the space and to obtain a_{ip}, the projections of points on an arbitrary set of orthogonal coordinate axes, because B^* can be written as follows:

$$B^* = AA' \qquad [7]$$

where $A = (a_{ip})$ and $p = 1, 2, \ldots, r$. A' is the transposition of A. It is important to note that Equation 7 does not hold unless B^* is positive semidefinite, that means, r latent roots of B^* have relatively large positive values and the remaining roots vanish except for errors. In other words, the colors will be considered as lying in a real Euclidean space of r dimensions if B^* is shown to be of rank r and to have no large negative latent root.

Two B^*s were constructed for Id and for K, O, and M. In accordance with Equation 7, B^* was factored by the centroid method in the case of Id and by the multiple-group method (Harman, 1954) in the other case. Three factors were extracted in both cases. For the following reasons, however, the third factor could be discarded and we could conclude that B^* was of rank two, except for variable errors, for Id as well as for K, O, and M: (*a*) The second factor residuals all seemed sufficiently small; (*b*) The contribution of the additional factor was relatively small. The ratio $\Sigma_i a^2_{i3}/\Sigma_i a^2_{i1}$ was .13 for Id and .17 for K, O, and M. (*c*) The criterion for B^* to be of rank two is given by Torgerson as follows:

$$\sum_i \sum_j (b^*_{ij})^2 = (\sum_i a^2_{i1})^2 + (\sum_i a^2_{i2})^2 + 2(\sum_i a_{i1}a_{i2})^2 \qquad [8]$$

In the case of Id, 81.18×10^7 was obtained for the left-hand term and 76.70×10^7 for the right. The coincidence seems close enough to discard the third factor, though it is a little worse than the coincidence in Torgerson's experiment (1958). The ratio between the left- and the right-hand members of the equation was 1 to .944 in the former and 1 to .981 in the latter. For the data of K, O, and M, the criterion would be inappropriate because of the method of factoring. In order to apply the criterion, it is desirable that the a_{ip}'s in the equation be the results of the method of principal axes or, at least, of the centroid method. (*d*) No particular relationship was found between a_{i3} and the fluctuation of the stimuli in Munsell value.

Results and Discussion

The configurations given by As are shown in Fig. 2 and Fig. 3 where the coordinate axes are omitted. Instead, radius vectors are drawn so as to run through colors of the same hue and a series of curves is also formed for chroma. Since the scale units have been equated as stated before, the two configurations can be directly compared with each other.

It is interesting to observe that

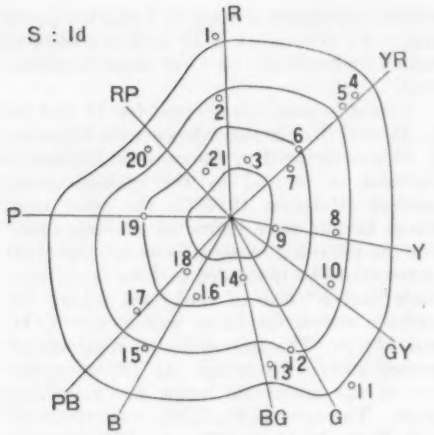

FIG. 2. Two-dimensional mapping of colors for Subject Id. Equichroma curves and radius vectors representing hue are formed in accordance with the configuration obtained.

the colors of the same Munsell hue appear along the line segments in the order corresponding to their chroma and, besides, all line segments converge into a single point. Furthermore, the pairs of the complementary

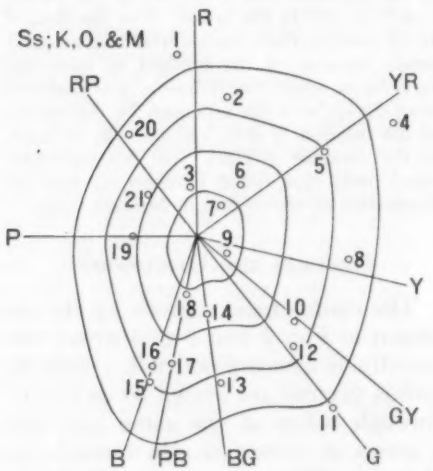

FIG. 3. Two-dimensional mapping of colors for *Ss* K, O, and M. Equichroma curves and radius vectors representing hue are formed in accordance with the configuration obtained.

hues are identified by the radius vectors lying more or less in the opposite direction to each other. Except for B and PB in Fig. 3, the order of hue is exactly what is expected. As to the spacing, however, the disagreement with the Munsell notation is conspicuous. Nevertheless, the patterns in Fig. 2 and Fig. 3 coincide closely with each other in the interval from one hue to the next. The exceptions are the position of G between GY and BG, and the positions of PB and B between BG and P. Comparison with Fig. 1 shows that the two spacings of hue obtained deviate more or less in the same way from the Munsell notation. For example, the interval between R and YR and also that between YR and Y are too large and the interval between B and PB is too small in the results of this study.

The equichroma curves in Fig. 2 and 3 were drawn so that, though the interval between the neighboring curves may differ from hue to hue, the interval on the same hue should be the same irrespective of chroma. If the series of curves in Fig. 2 and 3 are superimposed directly on each other, they are in good agreement in the region from RP through Y to BG. On the contrary, in the region of B and P the disagreement is marked. The disagreement, however, is apparently due to the discrepancy in spacing of B and PB. There is no difference between Fig. 2 and 3 of the interval from one equichroma curve to the next along each of these hues. It might be well to consider whether a set of concentric equichroma circles, as defined in the Munsell notation, can also be fitted to the data. Excepting the region around RP, the fitting was not entirely impossible and a plot of the chroma of the colors thus defined with the Munsell chroma of

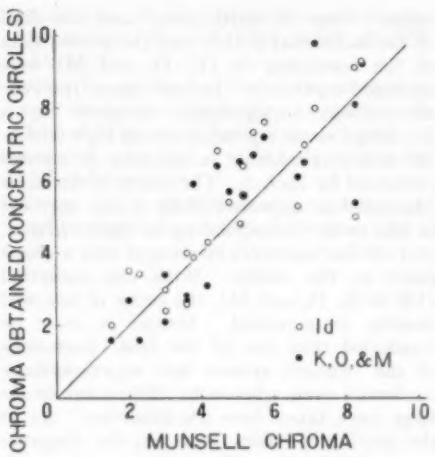

FIG. 4. Munsell chroma with chroma defined in terms of concentric equichroma circles fitted to the configurations in Fig. 2 and 3.

the stimuli is given in Fig. 4. The root mean square of the discrepancies between the two kinds of chroma was 1.24. If the restriction that equal intervals in Munsell chroma should correspond to equal linear extents irrespective of hue is withdrawn, obviously the goodness of fit improves. When the chroma in terms of the equichroma curves in Fig. 2 and 3 was taken as the ordinate, the scatter of the points was so reduced that the root mean square of the discrepancies decreased to 7.6.

Setting aside the question as to the adequacy of the spacing of the Munsell colors, it is satisfactory to observe that the colors varying in hue and chroma were fitted by the configuration in two-dimensional Euclidean space. If uniformity of the scale *in the large* is taken into consideration, among the basic postulates of the Munsell system the most questionable one will be that, of the polar coordinate in a Euclidean plane of constant lightness, hue is represented by the magnitude of the argument and chroma by the length of the radius vector. In order that the model may hold good, the difference perceived between two colors not of the same hue, say a and b, has to vary in a certain manner, if the chroma of each color is changed. Denoting by θ the difference between the two arguments corresponding to Hues a and b and by ρ_1 and ρ_2 the lengths of the two radius vectors representing the chromas of the pair of colors, the difference will be proportional to the distance

$$\sqrt{\rho_1^2 + \rho_2^2 - 2\rho_1\rho_2 \cos\theta} = \Delta$$

Let us consider, then, a pair of colors which are of Hues a and b respectively, but the chroma of each color is n times in magnitude that of the corresponding member of the former pair. By definition, the difference between the latter pair has to be proportional to the distance

$$\sqrt{n^2\rho_1^2 + n^2\rho_2^2 - 2n^2\rho_1\rho_2 \cos\theta} = n\Delta$$

That means, the difference between a pair of colors has to increase n times if both of the chromas of the members are increased n times. Otherwise, the model will not hold good.

It must be remembered that the configuration under discussion is more than a mere schema, because the points are located in such a way that the distances between all pairs of the points are proportional to the observed differences between the corresponding colors, no matter how large or small the differences are. Let s_{ij} denote the distance between the Points i and j in the multidimensional configuration. When the agreement between s_{ij}, i.e., the difference in theory, and d_{ij}, i.e., the observed difference scaled, was shown in terms of the correlation coefficient, r was .96 for Id. The multidimensional mapping for K, O, and M in Fig. 3 was made from the pooled data. Nevertheless, using s_{ij} determined from Fig. 3, r was .84 for K, .88 for O, and .85 for M. No qualitative individual differences could be observed among *S*s. However, the correlation coefficient will not be sufficient to tell the exact relationship existing between s_{ij} and d_{ij}. In the case of K, O, and M, as expected, the points plotted were scattered along the line passing the

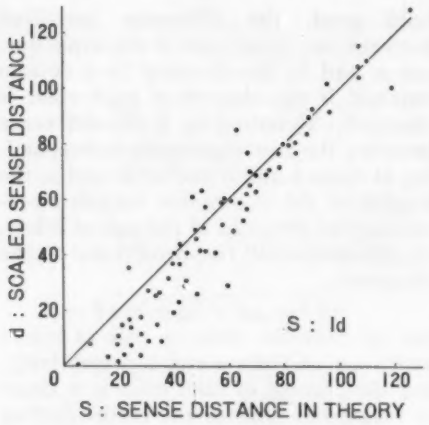

FIG. 5. Relation of theoretical sense-distances to scaled differences between all possible pairs of the colors. A third of all the pairs drawn randomly are plotted. Result for Subject Id. Theoretical sense-distance is taken from Fig. 2.

origin with the slope of one. As shown in Fig. 5 the regression of d_{ij} on s_{ij} is slightly curved, in the case of Id. The tendency that $d_{ij} < s_{ij}$ is predominant where $s_{ij} < 50$. It was mentioned previously that the distance $x_{ij(\delta)}$ may not be proportional to the perceived difference δ_{ij} between Colors i and j. Instead, it may be proportional to a function of δ_{ij}. The phenomenon observed in Fig. 5 may be ascribed partially to the non-proportionality, if any, existing between the observable x and the latent variable δ.

SUMMARY

The difference perceived between the standard color and each of 21 colors was converted into spatial distance by moving the latter one by one. As each of the 21 colors took the place of the standard, a matrix was obtained whose elements were the ratio scales of the differences between all possible pairs of the 21 colors. The adjustment was made with respect to the difference, if any, in the unit of converting, from trial to trial and from S to S. Then, the method of multidimensional scaling was applied to the matrix. The colors varied in hue and saturation over the widest possible range but were kept constant in lightness at 5 in Munsell value. Four Ss participated and the data of the individual S (Id) and the pooled data of the remaining Ss (K, O, and M) were analyzed separately. In both cases, the two-dimensional configuration obtained in a Euclidean space reproduced with high fidelity the structure inherent in the color differences perceived by each S. The colors of the same Munsell hue appeared along a line segment in the order corresponding to their chroma, and all line segments converged into a single point at the center. With one exception (PB in K, O, and M), the order of hue was exactly as expected. Hence, it may be concluded that one of the basic postulates of the Munsell system was experimentally confirmed even when color differences *in the large* were taken into consideration. As to the spacing of colors, however, the disagreement with the Munsell notation was observed, especially with respect to hue.

REFERENCES

GODLOVE, I. H. Improved color-difference formula, with applications to the perceptibility and acceptability of fadings. *J. Opt. Soc. Amer.*, 1951, **41**, 760–772.

HARMAN, H. H. The square root method and multiple group methods of factor analysis. *Psychometrika*, 1954, **19**, 39–55.

INDOW, T., & SHIOSE, T. An application of the method of multidimensional scaling to perception of similarity or difference in colors. *Jap. psychol. Res.*, 1956, **3**, 45–64.

JUDD, D. B. *Color in business, science, and industry*. New York: Wiley, 1952.

MESSICK, S. J. The perception of attitude relationships: A multidimensional scaling approach to the structuring of social attitudes. Unpublished doctoral dissertation, Princeton Univer., 1954.

MESSICK, S. J. An empirical evaluation of multidimensional successive intervals. *Psychometrika*, 1956, **21**, 367–375.

MESSICK, S. J., & ABELSON, R. P. The additive constant problem in multidimensional scaling. *Psychometrika*, 1956, **21**, 1–15.

NEWHALL, S. M. Preliminary report of the OSA subcommittee on the spacing of the Munsell colors. *J. Opt. Soc. Amer.*, 1940, **30**, 617–645.

NEWHALL, S. M., NICKERSON, D., & JUDD, D. B. Final report of the OSA subcommittee on the spacing of the Munsell colors. *J. Opt. Soc. Amer.*, 1943, **33**, 385–418.

RICHARDSON, M. W. Multidimensional psychophysics. *Psychol. Bull.*, 1938, **35**, 659. (Abstract)

SHIOSE, T. An application of the method of multidimensional scaling to the perception of similarity or difference in colors. II. *Jap. J. Psychol.*, 1958, **28**, 375–385. (English summary)

STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

TORGERSON, W. S. Multidimensional scaling: Theory and method. *Psychometrika*, 1952, **17**, 401–419.

TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

(Received June 8, 1959)

Journal of Experimental Psychology
Vol. 59, No. 5, 1960

MULTIDIMENSIONAL MAPPING OF MUNSELL COLORS VARYING IN HUE, CHROMA, AND VALUE[1]

TAROW INDOW AND KEI KANAZAWA

Keio University

Evidence in support of the basic postulates of the psychological color solid has been accumulated by the application of the multidimensional scaling. Torgerson (1958), Messick (1954), and Indow and Shiose (1956) showed that Munsell colors of the same hue (5R) could be fitted by a configuration in a two-dimensional Euclidean space, and that the orthogonal coordinates were identified as representing Munsell value (lightness) and chroma (saturation) respectively. Besides, the two-dimensional mapping of colors of constant lightness (Value 5) gave the following results (Indow & Uchizono, 1960). The colors lay in a Euclidean plane and each hue was represented by one of the radius vectors which converged into a single point corresponding to gray. Around this point, it was possible to define the series of equichroma curves. The hypothesis underlying the construction of an ideal color solid, e.g., the Munsell system, would be that equal sense distances shall correspond to equal linear extents in a Euclidean space, no matter how large or small the magnitude of the sense distances may be (Balinkin, 1939; Newhall, 1939). Since the multidimensional mapping is carried out on the basis of sense distances between every possible pair of colors employed, the findings mentioned above seem to provide experimental evidence in support of at least the fundamental hypothesis. As a matter of course, there are minor discrepancies between the results of the experiments and the Munsell spacing of color. The discrepancies will, however, be removed by adjusting the locations of the Munsell samples, if necessary. On the contrary, if the hypothesis were shown untenable, the color solid could not be more than a mere schematic representation.

In the previous experiments, the variation among colors has been more or less restricted. Colors have been used which differ from each other in only two attributes. Shiose (1958) made an experiment with the colors differing in value, chroma, and hue. However, the hue ranged only from 6R to 6YR. Messick (1956) reconsidered the results of his experiment taking into consideration slight hue variation of the colors used (roughly from 2R to 5.5R), and found that the variation could be accounted for by the addition of the third dimension. In the present experiment, the tridimensional mapping of colors over a considerable range was attempted. So far, no qualitative individual difference has been found in the multidimensional configuration constructed separately for each *S* (Indow & Shiose, 1956; Indow & Uchizono, 1960). Hence, the multidimensional scaling was applied to the pooled data of 4 *S*s participating in the

[1] This research was supported by a grant from Ministry of Education, Japan. The investigators are indebted to S. Ura. Under his supervision a part of the computations involved in this research was carried out by the electronic automatic computor in Japan Electronic Industry Development Association. Appreciation is also expressed to M. Matthewson for her helpful advice concerning English. Thanks are due to T. Uchizono for her help in the preparation of this report.

present study to give a common configuration.

METHOD

Apparatus and stimuli.—The viewing conditions and the manner of presenting the stimuli on a gray board were exactly the same as those in the previous experiment (Indow & Uchizono, 1960). The instructions given to *S* were also the same; namely, the perceived difference between a standard Color *i* and each of the Colors *j* ($j = 1, 2, \ldots, 24$) was to be converted into spatial distance by the method of multiple ratio judgment. Tristimulus values for the 24 colors under the given illumination were calculated from the spectral reflectances of the colored papers and the spectral energy distribution of the light source. Then, their equivalents in the recommended Munsell notation are obtained. The colors were of two levels in lightness. The brighter ones were from 4.3 to 4.9 and the darker ones were from 6.4 to 7.3 in terms of Munsell value. For convenience, the former will be called the stimuli of Value 5 and the latter the ones of Value 7 in the figures. A plot of the stimuli according to the hue and chroma appears in Fig. 1 where arbitrary radius vectors are given only for the sake of convenience. In other words, the hue names in the figure do not necessarily mean the corresponding principal hues in the Munsell notation.

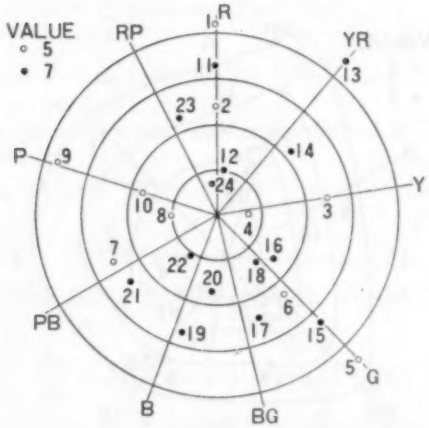

FIG. 1. Color stimuli in terms of Munsell hue and chroma. Concentric circles denote Chroma 2, 4, 6, and 8 respectively. Radius vectors are only for the sake of facilitating identification of colors.

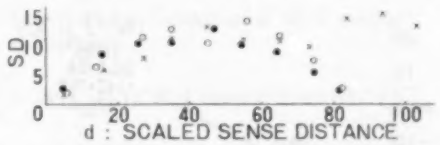

FIG. 2. *SD*s of scaled distances for four *S*s (×) and *SD*s of three repeated judgments for Subject Id (○) and for T (●) as functions of magnitude of the sense distance scaled.

Subjects.—All *S*s were normal in color vision. Id and Sh were investigators of the previous study (Indow & Shiose, 1956), while T and Si were students of Keio University who had neither experience in an experiment of this kind nor special knowledge of the Munsell system. Each *S* made three repetitions on each pair of Colors *i* and *j*.

Scaling procedure of sense distances.—As to the three spatial distances made by each *S* on the board to represent the perceived difference between the pair of Colors *i* and *j*, the average was obtained. To the set of the averages the scaling procedure was applied that was described in detail in the previous article (Indow & Uchizono, 1960). Thus, the sense distances were scaled once separately for each *S* to give $d'_{ij(\alpha)}$ where the subscript α denotes *S*. The next step was to equate e_α, the unit of the scale which may differ from *S* to *S*. Taking the unit of the scale of Id as standard, each of the remaining e_α's was multiplied by a constant so that a plot of the adjusted $d_{ij(\alpha)}$ against $d_{ij(\mathrm{Id})}$ was scattered along the line passing the origin with the slope of one. The scaled values expressed in terms of the common Unit e_{Id} were then averaged. The averages, d_{ij}'s, constituted a Common Distance Matrix *D* for the four *S*s. The multidimensional scaling procedures were applied to *D*.

The *SD*s of the four values combined, i.e., $d_{ij(\alpha)}$'s and $d_{ij(\mathrm{Id})}$, are plotted as a function of the magnitude of d_{ij} in Fig. 2. For the sake of comparison, the *SD*s of the three repeated judgments made by Id and also by T are shown in the same figure. The heterogeneity among *S*s was of the same magnitude as the fluctuations within the repeated responses of an individual *S*. It is interesting to observe that the errors are minimum at the lowest and at the highest values in the case of the individual data. This is the pattern that is often encountered in psychological tests (Mollenkopf, 1949).

Multidimensional scaling.—From Distance Matrix *D*, Matrix B^* was obtained which consists of b^*_{ij}, the scalar product of the vec-

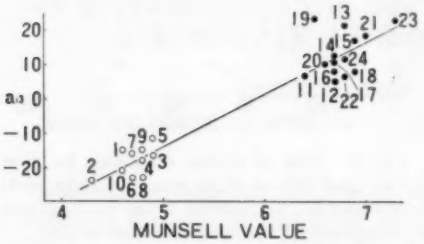

FIG. 3. Munsell value with projection of the points on third axis which is supposed to represent lightness of color.

tors to Points i and j from the centroid of the 24 stimuli used. Matrix B^* was then factored to give Matrix A, by means of obtaining the latent vectors and the latent roots of B^*. All the computations were carried out by the electronic automatic computor. Matrix A gave a mapping of the 24 colors in a Euclidean space of three dimensions. Matrix B^* was warranted of rank 3, since the third factor residuals were sufficiently small and the criterion given by Torgerson:

$$\sum_i \sum_j (b^*_{ij})^2 = \sum_p (\sum_i a^2_{ip})^2 + 2 \sum_{p \neq q} (\sum_i a_{ip} a_{iq})^2$$

was met when p was 3. The right- and left-hand members of the equation were 11.03×10^8 and 10.77×10^8, respectively. The ratio between the two is 1 to .976, whereas that in the study of Torgerson (1958) was 1 to .981. It was discovered that the fit to the Munsell notation became better when the coordinate axes were rotated. The projections of the points on each of the three orthogonal axes after the rotation are used as a_{ip}'s. As will be shown later, the first and second axes determine the plane of constant lightness, while the third corresponds to Munsell value.

RESULTS AND DISCUSSION

A plot of a_{i3}'s with the Munsell value of the stimuli is shown in Fig. 3. The separation of the two levels of lightness is perfect and, besides, the fluctuations in the Munsell value within each level is also somewhat reflected in the results, because the scatter within the level gives a correlation coefficient of .51 in the darker level and .40 in the brighter level.

The configuration in terms of hue and chroma is given in Fig. 4. The colors of the two lightness levels are plotted together. Radius vectors running through colors of the same hue are drawn and a series of curves is also formed for chroma on the same principle as in the previous study (Indow & Uchizono, 1960). The results are substantially in agreement with the results of the previous experiment. It is interesting that the configurations, in the brighter and darker levels, could be projected into a single plane without contradiction. For instance, YR in the brighter level comes between R and Y in the darker level, and Colors 15, 16, and 18 of the brighter level form a radius vector with 5 and 6 of the darker level in accordance with the definition of stimuli in Fig. 1.

In the previous study (Indow & Uchizono, 1960), two 2-dimensional configurations of a constant lightness were constituted, for individual data of Id and for the pooled data of K, O, and M, and no remarkable qualitative difference was found between the two. Hence, if the two configurations are combined into one, comparison between the results of the previous and present studies will be

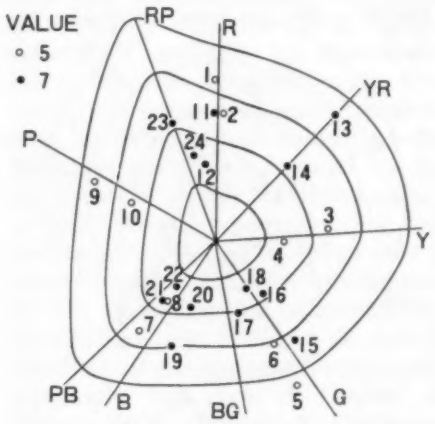

FIG. 4. Two-dimensional mapping of colors in terms of hue and chroma. Equichroma curves and radius vectors representing hue are formed in accordance with the configuration obtained.

facilitated. Combining two configurations is possible by the method of least squares.

Let us suppose that two configurations are given, both of which are represented in polar coordinates, as (ρ_{1i}, θ_{1i}) and (ρ_{2i}, θ_{2i}), where ρ denotes a radius vector and θ an argument with reference to an arbitrary axis. Then, the problems are to find the angle which the two reference axes should form, in order to give the closest agreement between the configurations, and also to find the coefficient π which should adjust the difference, if any, between the units of two scales. Denoting by α the angle between the reference axes of two coordinates, and by a_i the distance between a Point i in a configuration and its corresponding point in another, then

$$a_i^2 = \rho_{1i}^2 + \pi^2\rho_{2i}^2 - 2\pi\rho_{1i}\rho_{2i}\cos(\theta'_{2i} - \theta_{1i}) \quad [1]$$

where θ'_{2i} represents the argument of a Point i in the second configuration expressed in terms of the first polar coordinate, i.e.,

$$\theta'_{2i} = \theta_{2i} + \alpha$$

and the unit of the second is equated to that of the first.
Defining

$$A = \sum_i \rho_{1i}\rho_{2i}\cos(\theta_{2i} - \theta_{1i})$$

$$B = \sum_i \rho_{1i}\rho_{2i}\sin(\theta_{2i} - \theta_{1i})$$

the quantity which is to be minimized is given by

$$Q = \sum_i a_i^2 = \sum_i \rho_{1i}^2 + \pi^2 \sum_i \rho_{2i}^2 - 2\pi(A\cos\alpha - B\sin\alpha) \quad [2]$$

Differentiating Equation 2 with respect to α, we have

$$\frac{\partial Q}{\partial \alpha} = 2\pi(A\sin\alpha + B\cos\alpha) \quad [3]$$

The minimizing value $\hat{\alpha}$ is found by equating Equation 3 to zero. Simplifying the result gives

$$\hat{\alpha} = \tan^{-1}\frac{-B}{A} \quad [4]$$

Differentiating Equation 2 with respect to π, we have

$$\frac{\partial Q}{\partial \pi} = 2\pi\sum_i \rho_{2i}^2 - 2(A\cos\alpha - B\sin\alpha) \quad [5]$$

The minimizing value $\hat{\pi}$ is found by equating Equation 5 to zero. By using Equation 4, we have

$$\hat{\pi} = \frac{1}{\sum_i \rho_{2i}^2}(A\cos\hat{\alpha} - B\sin\hat{\alpha}) \quad [6]$$

The two-dimensional configuration of K, O, and M (Indow & Uchizono, 1960) was superimposed upon that of Id (Indow & Uchizono, 1960) in this way ($\hat{\pi} = .94$, $\hat{\alpha} = 3°\ 30'$, taking R as the reference axis). Taking the middle point of a_i, the combined configuration of the colors was developed, which is shown in Fig. 5 where the numbering of the colors is to be understood as defined in Fig. 1 in Indow and Uchizono (1960). Radius vectors and a series of equichroma curves are formed on the same principles as before. Comparison with Fig. 4 shows that agreement between the results of the two experi-

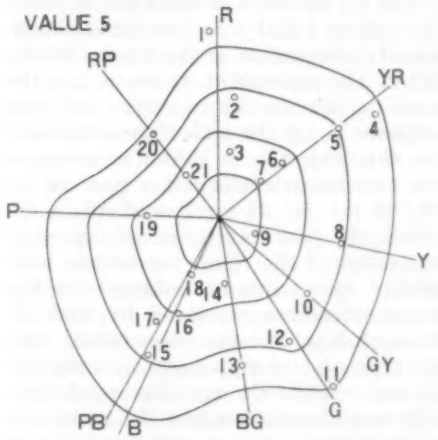

FIG. 5. Two configurations obtained by Indow and Uchizono (1960) are combined into one in order to be compared with Fig. 4. Number of colors corresponds to that given in Fig. 1 of Indow and Uchizono. Equichroma curves and radius vectors representing hue are formed in accordance with the configuration shown here.

ments is satisfactory, except for the definition of chroma in the region of RP and the spacings of P and RP. As a matter of course, Fig. 5 can be superimposed upon Fig. 4 by the method described above. In so doing, it may be appropriate to take as references the points of intersection of the equichroma curves with the radius vectors used in common in both investigations. Taking R as the reference axis, it was found that the closest fit was obtained by $\hat{\pi} = 1.02$ and $\hat{\alpha} = -2° 52'$. Should loci of constant hue and constant chroma be defined as averages of the results of the two experiments, the equichroma curves do not deviate so much from the concentric circles, and the loci of constant hue are substantially in agreement with the Munsell spacings except for PB. Radius Vector PB appeared too close to B in both of the two experiments. It is the impression of the investigators that, looking at the Munsell Book of Color (Pocket Edition), people seem unanimous in pointing out that the size of the step from one sheet of constant hue to the next is relatively small in the region from B to PB.

Let s_{ij} denote the distance between the points i and j in the three-dimensional configuration of the present study. When the agreement between s_{ij}, the color difference in theory, and the adjusted $d_{ij(\alpha)}$, the scaled sense distance for the subject α, is shown in terms of the correlation coefficient, r was .94 for Id, .90 for Sh, .94 for T, and .95 for Si. Obviously the multidimensional configuration of the colors reproduces with fidelity the structure inherent in the color differences perceived by each *S*. Nevertheless, it was discovered that the regression of $d_{ij(\alpha)}$ upon s_{ij} is slightly curved. While the curvilinear relationship was observed only in Id among *S*s participating in the previous experiment (Indow & Uchizono, 1960), it was conspicuous in all of the four *S*s in the present experiment. The relationship observed in Id in this experiment was exactly the same as that in the previous one. In Fig. 6 is given a typical example of the relationship observed in the

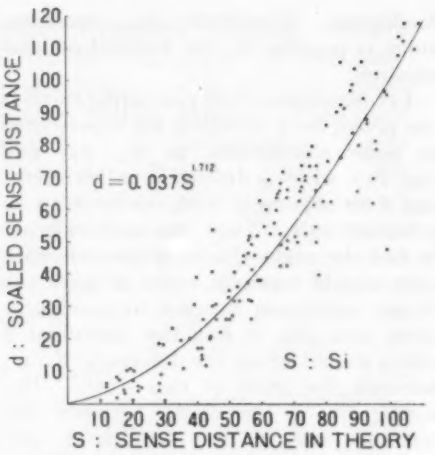

FIG. 6. Relation of theoretical sense distances to scaled differences between all possible pairs of the colors. A third of all the pairs drawn randomly are plotted. Result for Subject Si. Theoretical sense distance is taken from Fig. 3 and 4.

remaining *S*s that seems to be more or less parabolic in all the cases. The empirical equation in Fig. 6 was fitted by inspection. The phenomenon seems to be a subject eminently worthy of theoretical study.

When the chroma of the colors defined in terms of the equichroma curves in Fig. 4 was plotted against the chroma in the Munsell notation of the stimuli, it became apparent that the fit was not satisfactory. The root mean square of the discrepancies was .97 and this was not so large compared with .76 in the previous study. Nevertheless, the tendency was apparent that the points systematically deviated above the line in the region of the lower chroma. Hence, the size of the step from Chroma 0 to 2 was defined differently from the size of the other steps along the same vector that was kept constant. The equichroma curves thus defined are given in Fig. 7. When the chroma in terms of these curves was plotted against the chroma in the Munsell notation, the above mentioned tendency disappeared and the goodness of fit improved as shown by the fact that the root mean square of the discrepancies was .73. It is to be

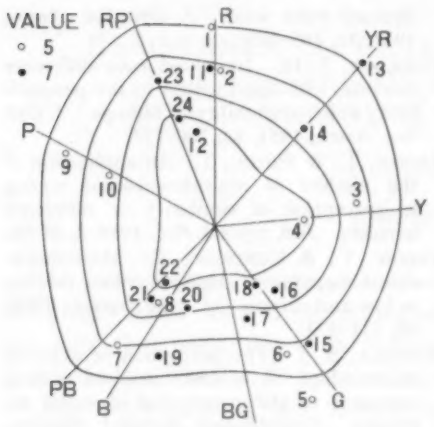

FIG. 7. Two-dimensional mapping of colors where equichroma curves are formed on the principle that size of the step from Chroma 0 to 2 may differ from sizes of the other steps which are kept constant in each hue.

noted that the equichroma curves approximate concentric circles on this principle. The loci of Chroma 2 were irregular in shape. Bellamy and Newhall (1942) determined limens with respect to chroma from achromatic origin in the five principal hues. The limens were given in terms of the Munsell scales. No particular relation was found, however, between the size of the first chroma step in the present experiment and that of the stimulus limen of chroma in the experiment of Bellamy and Newhall. It is most likely that there is a causal connection between this phenomenon, i.e., that the first chroma step was wider, and the curvilinear relationship between s and d. If the multidimensional scaling is applied to the absolute interstimulus distances which are obtained by adding a constant to relative interstimulus distances, the choice of a particular value of the additive constant will affect in theory the subsequent multidimensional mapping of the stimuli. Messick and Abelson (1956) showed that, if the additive constant is overestimated in a two-dimensional mapping, the configuration of stimuli becomes larger and "convex," and if underestimated, the configuration becomes smaller and "concave." It is true that, in the present study, the color differences have been obtained on a ratio scale from the beginning to make the estimation of the additive constant unnecessary, and that no curvature was discovered in the planes of constant lightness when a_{i3} was plotted against a_{i1}, or a_{i2}. We have no hypothesis at present to account for the problem under discussion. Certainly it awaits further inquiry.

In the Munsell system, the relative size of the units of the three attributes has been in dispute. For instance, Nickerson and Stultz (1944) consider that 1, 2, and 3 steps in V, C, and H respectively are visually equivalent in the region of 5C, and 1, 1, and 2 steps in the region of 3C, whereas, according to Bellamy and Newhall (1942), 1, 8, and 22 steps in V, C, and H respectively are defined as the same in perception in the region of 6C. On the other hand, 1 step in V is said to correspond to 4 steps in C by Godlove (1951). In the two-dimensional configurations individually constructed, Indow and Shiose (1951) obtained the following correspondences between the steps in V and in C; 1 to 3.1, 1 to 3.0, and 1 to 2.3 for the three *Ss* respectively. The correspondence in the present study was 1 to 1.8, when the average of the sizes of the one-chroma steps in all the hues in Fig. 4 was compared with the size of the one-value step in Fig. 3. From Fig. 3 and 5, it may also be inferred that the correspondence was 1 to 2.0 in the average in the experiment by Indow and Uchizono (1960). In any case, the relative size of chroma step appears larger in these two experiments than in the previous experiment by Indow and Shiose. It is to be noted, however, that the colors used in the experiment by Indow and Shiose ranged in chroma from 4 to 12. If the correspondence is determined in the present experiment from Fig. 3 and 7, that is to say, if the sizes of the steps above Chroma 2 only are averaged, 1 step in V becomes equivalent to 2.7 steps in C, which is in agreement with the correspondence obtained in the experiment by Indow and Shiose.

Summary

The sense distances between all possible pairs of the 24 colors were scaled by the method of multiple ratio judgment and by the procedures of adjusting the differences, if any, in the unit by which *Ss* converted color differences into spatial distances. Then, the method of multidimensional scaling was applied. As the colors varied in the three attributes, the three-dimensional configuration of the colors was obtained for the pooled data of the four *Ss* participating. The agreement among *Ss* was satisfactory in spite of the fact that half of the *Ss* were inexperienced. The axis corresponding to the Munsell value was orthogonal to the plane of constant lightness and the configuration of the colors in terms of Munsell hue and chroma was substantially in agreement with that obtained in the previous study. Both deviated somewhat in the same way from the Munsell notation, especially in the spacing of hue. It was pointed out that the equichroma curves fit the data better if the size of the chroma step from 0 to 2 is allowed to be wider than that of the remaining chroma steps. The correspondence between the steps of Munsell value and of Munsell chroma was 1 to 1.8 when the equichroma curves were drawn evenly in each hue. It becomes 1 to 2.7, however, with respect to the chroma steps, except for the first ones, if the latter are made wider than the former. It was discovered that the relation of the distances between all possible pairs of the points plotted in the three-dimensional configuration to the scaled sense distances from which the mapping was made was not linear but rather parabolic.

References

Balinkin, I. A. Industrial color tolerances. *Amer. J. Psychol.*, 1939, **52**, 428–448.

Bellamy, B. R., & Newhall, S. M. Attributive limens in selected regions of the Munsell color solid. *J. Opt. Soc. Amer.*, 1942, **32**, 465–473.

Godlove, I. H. Improved color-difference formula, with applications to the perceptibility and acceptability of fadings. *J. Opt. Soc. Amer.*, 1951, **41**, 760–772.

Indow, T., & Shiose, T. An application of the method of multidimensional scaling to perception of similarity or difference in colors. *Jap. psychol. Res.*, 1956, **3**, 45–64.

Indow, T., & Uchizono, T. Multidimensional mapping of Munsell colors varying in hue and chroma. *J. exp. Psychol.*, 1960, **59**, 321–329.

Messick, S. J. The perception of attitude relationships: A multidimensional scaling approach to the structuring of social attitudes. Unpublished doctoral dissertation, Princeton Univer., 1954.

Messick, S. J. An empirical evaluation of multidimensional successive intervals. *Psychometrika*, 1956, **21**, 367–375.

Messick, S. J., & Abelson, R. P. The additive constant problem in multidimensional scaling. *Psychometrika*, 1956, **21**, 1–15.

Mollenkopf, W. G. Variation of the standard error of measurement. *Psychometrika*, 1949, **14**, 189–229.

Newhall, S. M. The ratio method in the review of the Munsell colors. *Amer. J. Psychol.*, 1939, **52**, 394–405.

Nickerson, D., & Stultz, K. F. Color tolerance specification. *J. Opt. Soc. Amer.*, 1944, **34**, 550–570.

Shiose, T. An application of the method of multidimensional scaling to perception of similarity or difference in colors. II. *Jap. J. Psychol.*, 1958, **28**, 375–385. (English summary).

Torgerson, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

(Received June 8, 1959)

Journal of Experimental Psychology
Vol. 59, No. 5, 1960

RESISTANCE TO EXTINCTION AS A FUNCTION OF DEGREE OF REPRODUCTION OF TRAINING CONDITIONS[1]

MELVIN H. MARX

University of Missouri

The present study is one of a series designed to test some implications of the view that resistance to extinction is primarily a function of maintenance of the motivation of the animal to continue to perform the instrumental act, independently of any variation of the drive factor manipulated by incentive deprivation. In a previously reported experiment (Marx, 1958), animals extinguished with an empty foodcup present on half of the trials showed greater resistance to extinction than animals run with an empty foodcup present on every trial, following continuous reward during training.

The particular hypothesis tested in this experiment is that resistance to extinction is positively related to the degree of reproduction of training conditions. Hungry animals were trained in a runway with three distinctive endboxes, only one of which ever contained food. They were extinguished with only the formerly positive, only one of the formerly negative, both of the formerly negative, or all three of the endboxes present.

[1] This investigation was supported in part by a research grant, M-817, from the National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland. Thanks are due to Donald Morgan, who collected the data, and Aaron J. Brownstein, who performed the statistical computations and criticized the manuscript.

METHOD

Subjects

The *S*s were 36 experimentally naive female albino rats, approximately 120 days old at the start of the experiment.

Apparatus

A simple straight runway was used. It had a 1-ft. startbox, a 7-ft. stem, and a 1½-ft. endbox (see Marx, 1958, for further details). Starting time, running time, and time in the endbox were all recorded automatically by means of floor switches. Black, white, and striped endboxes were used in training; a neutral Masonite endbox was used in accommodation.

Experimental Design

All *S*s were given the same accommodation and training. White, black, and striped endboxes were used in training, but each *S* found food only in one of these. Three *S*s in each group found food in each of the three endboxes. Extinction was run under four different endbox conditions (nine *S*s per group): Group P: formerly positive only; Group N: one of the formerly negative endboxes only; Group NN: both of the formerly negative endboxes; Group NNP: both of the formerly negative endboxes and the formerly positive endbox.

Procedure

Accommodation.—The *S*s were placed on a 23-hr. food deprivation schedule for five days prior to the start of accommodation to the apparatus. They were fed Little Friskie pellets for 1 hr. per day.

On the first day of accommodation, *S* was placed in the apparatus with all doors open, the neutral insert in the endbox, and no foodcup present. It was allowed to explore the runway and endbox for 15 min. A foodcup with five Little Friskie pellets was placed in the endbox and *S* was permitted

to eat these. Then *S* was removed from the apparatus and placed in an individual detention cage.

After 1 hr. *S* was returned to the apparatus and placed in the endbox by the foodcup, which contained two pellets. It was allowed to eat these and was then returned to the detention cage. After 40 sec. it was returned to the endbox just inside the closed door, at the end opposite to the foodcup. After *S* ate the two pellets in the cup, it was returned to the detention cage. Five additional trials were given at intervals of 40 sec. in the same manner as described above except that *S* was placed progressively further away from the endbox. The final trial was begun in the startbox, with all doors open. Following this trial *S* was placed in the detention cage for 5 min., after which it was returned to its home cage and fed for 1 hr. On the second day of accommodation the essential procedure of the first day was repeated, except for the initial 15-min. period and except that two trials were run from each starting position (to a total of 14 trials).

Training.—The *S*s were started in squads of four. They were given 12 training trials per day for 6 days. Each endbox was used four times, on a random schedule, in every set of 12 trials. Two pellets (weighing approximately .10 gm.) were in the foodcup in one of the endboxes only. The *S* was placed in the detention cage for 20 sec. between trials.

On the first training day an initial trial from the startbox was given with the neutral endbox in place and with all doors open, prior to the regular 12 trials with the startdoor closed. On each trial *S* was placed in the startbox and the startdoor opened 3 to 5 sec. thereafter. If *S* did not complete the run to the endbox within 180 sec., it was gently guided down the runway to the endbox floor switch. It was allowed 10 sec. in the endbox; on rewarded trials, if it had not eaten in that time, it was placed by the foodcup and given an additional 10 sec. to eat. At the end of the 12 trials *S* was placed in the individual detention cage for 5 min., after which it was returned to the home cage. When all four *S*s in a squad were back in the home cage, they were fed for 1 hr.

Extinction.—The *S*s were placed in the four groups randomly, within squads. Massed extinction trials were given, at an interval of 20 sec., to a criterion of two successive failures to enter the endbox with a minimum of 24 trials. Of the nine *S*s in each group, three had been run to each type of endbox under reward conditions. Group P now found only the formerly positive endbox, Group N only one of the formerly negative endboxes, Group NN both of the formerly negative endboxes (in the same order in which they had occurred during training), and Group NNP all three endboxes (in the same order in which they had occurred during training).

RESULTS

Training

By the sixth day of training essentially the same nearly asymptotic starting and running times had been reached by all four groups. In starting time, for example, all of the group medians for the 12 trials were within the range of 1.0 to 1.1 sec. In running time, three of the medians were 1.6 and one was 1.7 sec. No difference among groups was apparent in rate of approach to these values.

Time required to traverse the 18 in. between floor switches in the endbox is a somewhat more interesting measure because it provides a direct index of *S*'s discrimination between positive and negative goal conditions. Here again the groups reached essentially similar and nearly asymptotic running times. A clear and increasing difference between goal times in the positive and negative endboxes occurred from the second day on. With all 36 animals considered, median daily times were as follows: for the positive endbox, 7.3, 4.0, 2.3, 1.6, 1.5, and 1.3 sec; for the two negative endboxes, 8.0, 5.5, 5.0, 4.6, 4.8, and 4.9 sec. On the last day of training, each of the 36 *S*s ran faster in the positive endbox than in the negative endboxes.

Extinction

Trials to extinction.—Increasing resistance to extinction was found in the direction predicted. Trials to an extinction criterion of two successive failures to complete the

TABLE 1

MEAN NUMBER OF TRIALS TO REACH GOAL TIME CRITERION 1 (FIRST 10-SEC. SCORE) AND GOAL TIME CRITERION 2 (TWO SUCCESSIVE 10-SEC. SCORES, WITH CRITERION 1 SCORES EXCLUDED)

Criterion	Group			
	N	P	NN	NNP
1	3.1	10.7	4.0	4.8
2	8.0	3.9	4.1	15.6

running response within 90 sec. gave the following mean values: Group P, 45.7; Group N, 50.7; Group NN, 60.3; and Group NNP, 94.3. This difference among groups was highly reliable ($F = 12.11$, $P < .001$ for df 3/32). Direct comparisons of pairs of groups indicate that only the difference between Group NNP and the other groups was reliable (mean difference of 20.5 required for $P = .05$).

Trials to a less stringent criterion of extinction, the first failure to respond, gave essentially the same group differences but at a somewhat lower level of reliability ($F = 4.53$, $P < .01$, for df 3/32).

Group comparisons were also made for extinction of the running response within the goalbox. Two criteria were used: (*a*) number of trials to the first 10-sec. goal time; and (*b*) number of trials to two successive 10-sec. goal times. Here the influence of the formerly positive endbox is more clearly shown, as would be expected on the basis of the discrimination evident in training. The mean numbers of trials to meet each criterion are shown for the four groups in Table 1. Group P required more trials than Group N ($P < .05$) and than Groups NN and NNP ($P < .10$) for Criterion 1 and Group NNP reliably more than all other groups ($P < .05$) for Criterion 2.

A direct comparison of positive endbox times with negative endbox times within Group NNP gave the following results: 8.4 mean trials to the first 10-sec. trial for the positive, 3.4 for the negative; 10.3 additional mean trials to two successive 10-sec. trials for the positive, 8.8 additional for the negative.

Starting and running times.—The increments in starting and running times showed the same intergroup relations as were found for the trials-to-extinction measure. Interpretation of these gross data is somewhat difficult, however, because of the random admixture of endbox conditions in Group NN and Group NNP.

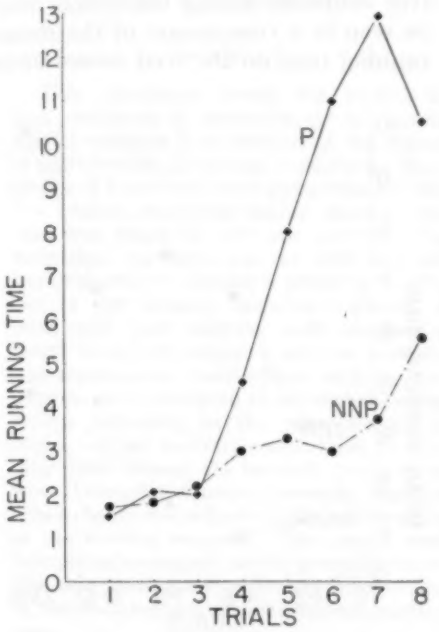

FIG. 1. Mean running times following presentation of previously positive endbox during extinction for Group NNP and comparable trials for Group P.

TABLE 2

ANALYSIS OF VARIANCE ON RUNNING TIME FOLLOWING PRESENTATION OF PREVIOUSLY POSITIVE ENDBOX DURING EXTINCTION FOR GROUP NNP AND COMPARABLE TRIALS FOR GROUP P[a]

Source	*df*	*MS*	*F*
Groups	1	399.03	8.82*
Error	14	48.44	
Trials	7	127.19	6.65**
G × T	7	56.18	2.94**
Error	98	19.12	

* $P < .05$.
** $P < .01$.
[a] One *S* in Group P had a criterion trial too early to be used in this analysis; to keep *N*s equal, one *S* in Group NNP was randomly discarded. If these *S*s were used, giving the Group P *S* an arbitrary score, differences would be magnified.

For this reason more analytic running-time comparisons are reported.

The effect of interpolation of negative endboxes during extinction may be seen in a comparison of the mean running time on the trial *immediately following* entrance into a formerly positive endbox (Group NNP) with that on comparable trials where only the previously positive endbox is presented (Group P). These data are shown graphically in Fig. 1. The clearly superior maintenance of running speed in Group NNP was highly reliable, as is shown in the statistical analysis presented in Table 2.

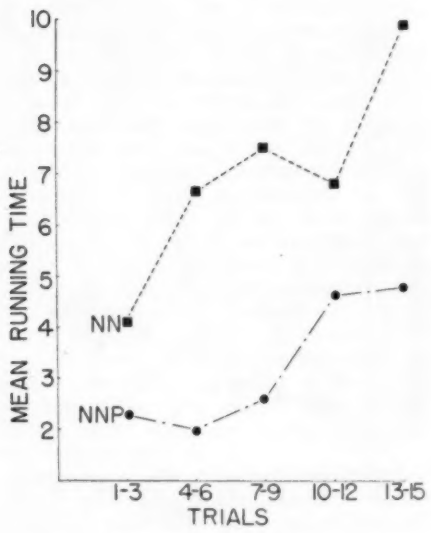

FIG. 2 Mean running times following presentation of previously negative endboxes during extinction for Group NNP and comparable trials for Group NN.

TABLE 3

ANALYSIS OF VARIANCE ON RUNNING TIME FOLLOWING PRESENTATION OF PREVIOUSLY NEGATIVE ENDBOXES DURING EXTINCTION FOR GROUP NNP AND COMPARABLE TRIALS FOR GROUP NN

Source	*df*	*MS*	*F*
Groups	1	936.32	4.806*
Error	16	194.81	
Trials	14	80.73	<1
G × T	14	64.23	<1
Error	224	98.09	

* $P < .05$.

The effect of interpolation of a positive endbox may be seen in a comparison of the mean running time on the trial immediately following entrance into a formerly negative endbox (Group NNP) with that on comparable trials where only the two formerly negative endboxes are presented (Group NN). These data are shown in Fig. 2, for pooled blocks of three trials, in order to show the trends more clearly. Here the group difference, but not the Group × Trials interaction, was reliable, as shown in Table 3. The mean running times were 3.25 sec. for Group NNP and 6.98 sec. for Group NN.

No reliable differences were found in running time between comparable trials in Groups N and NN ($F < 1.00$). No important effect can be attributed to the interpolation of a single

formerly negative endbox in the absence of any positive endboxes.

Discussion

The results of the experiment clearly support the prediction that resistance to extinction is a function of the degree of reproduction of the training conditions.

These data are consistent with the discrimination hypothesis (Bitterman, Fedderson, & Tyler, 1953; Elam, Tyler, & Bitterman, 1954). However, the failure of the simple N condition to show reliably greater resistance to extinction than the simple P condition does not support this interpretation. Also, the discrimination hypothesis was contraindicated by the results of an experiment previously reported (Marx, 1958). A recent experiment by Brown and Bass (1958) may be cited as failing to substantiate a discrimination or perceptual-change interpretation of extinction, while pointing to the importance of the variety of stimulus conditions as a strong contributor to resistance to extinction.

The most interesting aspect of the present data concerns the analytic running-time measures shown graphically in the figures. The main conclusion to be drawn from these comparisons is that the positive endbox, when interpolated among the negative endboxes on successive trials, is the critical factor in the situation. The statistically reliable differences found when Group NNP is compared with Group P and Group NN are to be contrasted with the unreliable difference found in the comparison of Groups N and NN. The importance of the interpolation of negative cues in extinction is also indicated by the comparison of goal time extinction scores for Groups P and NNP (see Table 1); Group P ran longer to the first, but Group NNP longer to the second consecutive, failure to enter the endbox.

According to the present interpretation the formerly positive endbox acts as a strong motivator when interpolated among formerly negative endboxes. Apparently the contrast serves to maintain *S*'s instrumental motivation in the same way as the cup–no cup condition did in the previous experiment.

Theoretical interest attaches to the fact that the differential effect was obtained by "interpolation" of the formerly positive endbox, as compared with a formerly negative endbox, when running speed following presentation of comparable negative endboxes was measured (Fig. 2). This fact suggests that a simple interpretation of these over-all differences solely in terms of the differential extinction of the secondary reinforcing properties of a positive cue is inadequate. The effect appears to be a more general motivational one, since it apparently carries over to running responses following negative cues, although the interpolated presentation of the positive cue as the focal point of the effect confirms the results of the previous experiment (Marx, 1958).

Summary

This experiment tested the hypothesis that resistance to extinction of an instrumental response is a function of the degree of reproduction of training conditions. Four groups of 9 rats each were given three distinctive endbox conditions during training, but food was found in only one of these. In extinction, one group was run with only the formerly positive endbox, a second with only one of the formerly negative endboxes, a third with both negative endboxes, and a fourth with both negative and the positive. The results were in accordance with the prediction, with resistance to extinction among groups increasing in the order mentioned above. More analytic comparisons of running times showed the formerly positive cue when interpolated among formerly negative cues to be the major factor in the maintenance of the running response. The results were interpreted as support for the general proposition that resistance to extinction is a function of the maintenance of instrumental motivation.

References

Bitterman, M. E., Fedderson, W. E., & Tyler, D. W. Secondary reinforcement

and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, **66**, 456–464.

BROWN, J. S., & BASS, B. The acquisition and extinction of an instrumental response under constant and variable stimulus conditions. *J. comp. physiol. Psychol.*, 1958, **51**, 499–504

ELAM, C. B., TYLER, D. W., & BITTERMAN, M. E. A further study of secondary reinforcement and the discrimination hypothesis. *J. comp. physiol. Psychol.*, 1954, **47**, 381–384.

MARX, M. H. Resistance to extinction as a function of continuous or intermittent presentation of a training cue. *J. exp. Psychol.*, 1958, **56**, 251–255.

(Received June 22, 1959)

Journal of Experimental Psychology
Vol. 59, No. 5, 1960

REPLICATION REPORT: AN ATTEMPT TO OBTAIN INHIBITION WITH REINFORCEMENT[1]

STEPHANIE S. FUCHS[2]

Reed College

Kendrick (1958) has reported the experimental validation of a deduction from Hullian theory which predicts the extinction of a continuously reinforced response. The present study is a replication of his experiment with respect to the pre-experimental handling of *S*s, the 30 trials per day, the criterion of extinction, the 20-hr. thirst drive, the ½-cc. water reward, and the other details described in his article (Kendrick, 1958, p. 314) except as follows: Kendrick's *S*s were chosen from populations characterized by extremes of emotionality; *S*s in the present study were six male Sprague-Dawley albino rats, 118 days old at the beginning of the experiment. The *S*s were run one at a time, in the middle of a 9-hr. dark period; Kendrick's *S*s were run two at a time, after the first 6 hr. of a 17½-hr. light period.[3] As in the Kendrick study, Days 1–3 were training days during which *S*s explored the alley; Day 4 was the first day of testing.

No extinction of the running response occurred in the present study within 41 days of 30 trials per day. In comparison, the mean number of days to extinction for Kendrick's *S*s was 32 days, and his longest-running *S*s extinguished on Day 41. Thus the Kendrick results were not confirmed.

Figure 1, showing median running times for the group for the testing days, indicates that running time was at asymptote for over 25 days. Latency results were very similar and less variable from day to day. Figure 1 also indicates that running times were very consistent on the first few trials of each day throughout the testing series, while, in contrast to this, on the last six trials, running times were extremely variable and much longer. Figure 2, showing intraday variation in running time, indicates that running times for the last six trials were consistently longer than those for the first six.

The behavior of *S*s was recorded for all trials; analysis of these data for those trials which were longer than 4 sec. indicates that *S*s are more active during the long later trials than during the long trials at the beginning of a day. During the early long trials, *S*s tended to walk slowly down the alley, sniffing

[1] This report is based upon a thesis submitted to the Division of Philosophy, Psychology, and Education of Reed College, in partial fulfillment of the requirements for the B.A. degree. The author is indebted to R. E. Boyle and R. M. Hanson for advice and encouragement during the course of this study.

[2] Now at University of Washington, Seattle.

[3] D. C. Kendrick. Personal communication. 1959.

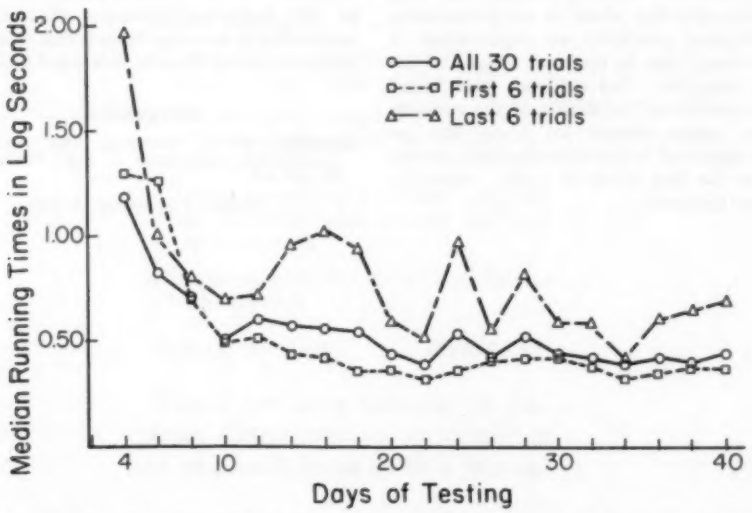

FIG. 1. Performance curves based on the median of the six median running times for all trials, for the first six trials, and for the last six trials per day.

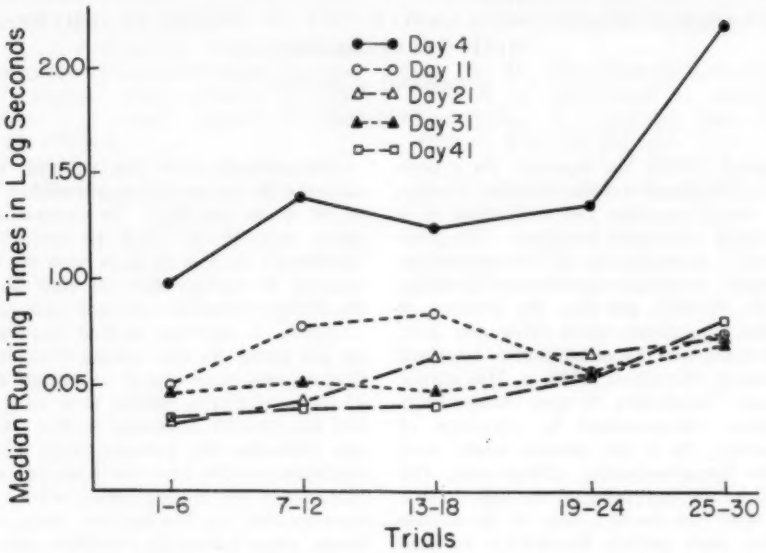

FIG. 2. Performance curves based on the median running times for successive fifths of the day's testing on selected days.

its floor and sides. This behavior rarely occurred at the end of a day's testing and is in marked contrast to the typical behavior during long late trials, which included such activity as running on the edge of the alley and retracing several times in the alley; these latter types of behavior almost never occurred in the early long trials of a day.

The increased activity during the long later trials of a day when *Ss* are presumably more fatigued precludes an explanation of long running times in terms of fatigue of the running response. No evidence was found for the conditioned inhibition of the running behavior, since almost no variability or increase appeared in the latencies and running times on the first trials of a day, when *Ss* were least fatigued.

It is possible that extinction would have occurred had *Ss* been run for some longer period. Factors which might affect the length of time necessary to achieve extinction are the strain of *Ss*, the amount of emotionality, and possible differences in activity level due to the differences in the relationship of the time of testing to the light/dark cycle. However, in the present study, the analysis of *Ss*' behavior indicates that increased variability in running times is due to increased activity, rather than to increased fatigue.

REFERENCE

KENDRICK, D. C. Inhibition with reinforcement (conditioned inhibition). *J. exp. Psychol.*, 1958, **56**, 313–318.

(Received November 16, 1959)

Wiley BOOKS

Ready in August—

EXPERIMENTAL PSYCHOLOGY

By BURTON G. ANDREAS,
Associate Professor of Psychology,
University of Rochester.

A new text which promotes student growth toward research competence . . .

Andreas' *Experimental Psychology* not only discusses the general principles and methods of experimentations, it also presents important recent experiments which illustrate the application of these principles and methods. Since these experiments can be easily modified for duplication in the undergraduate lab, recent research, text material, and lab work can be related in such a way as to give the student a feeling of participation in an ongoing scientific enterprise. To aid the student in his experimentation, sections on statistical inference and report-writing are included in the book.

Other features . . .

- Offers broad coverage of the traditional subject areas of the field
- Can be easily adapted to the needs and requirements of individual instructors
- Begins by giving a general orientation to behavioral science and experimentation
- Suggests possible directions for future research

Approx. 544 pages. *Prob. $6.95.*

Watch for more features of Andreas' *Experimental Psychology* in the next issue of this journal

JOHN WILEY & SONS, Inc.
440 Park Avenue South
New York 16, N.Y.

Mathematical Methods in the Social Sciences, 1959

PROCEEDINGS OF THE FIRST STANFORD SYMPOSIUM

EDITED BY KENNETH J. ARROW, SAMUEL KARLIN, PATRICK SUPPES

In June, 1959, twenty-five men in the fields of econometrics, operations research, and mathematical learning theory met at Stanford University to exchange ideas on the applications of mathematics to the social and managerial sciences. These are the papers presented at the conference, arranged in three sections: Economics, Management Science, and Psychology and Sociology. The contributors are: Gerard Debreu, Leonid Hurwicz, Lionel McKenzie, Michio Morishima, Paul Samuelson, Robert Solow, H. Theil, and T. Kloek, Hirofumi Uzawa, George Dantzig, Augustus Fabens, David Kendall, T. W. Anderson, Richard Atkinson, C. J. Burke, Robert Bush, W. K. Estes, Eugene Galanter, George Miller, Duncan Luce, Jacob Marschak, Herbert Solomon, Patrick Suppes, and the editors. *Mathematical Studies in the Social Sciences, IV.* *$8.50*

STANFORD UNIVERSITY PRESS

Order from your bookstore, please